DOCUMENT RESUME

ED 187 741 TM 800 237

AUTHOR Haney, Walt; Radin, Eugene
TITLE Summary of the Fall 1979 Conference of the National Consortium on Testing; November 5-6, 1979.
INSTITUTION Huron Inst., Cambridge, Mass.
PUB DATE 1 Dec 79
NOTE 104p.; For related documents, see TM 800 235-236.
AVAILABLE FROM Huron Institute, National Consortium on Testing, 123 Mt. Auburn Street, Cambridge, MA 02138 ($5.00)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *College Entrance Examinations; *Court Litigation; Curriculum; Educational Legislation; *Educational Testing; Information Utilization; Intelligence Tests; Legal Problems; Minimum Competency Testing; Professional Associations; School Districts; State Programs; Test Coaching; *Testing Problems; *Testing Programs
IDENTIFIERS California; Debra P v Turlington; Federal Trade Commission; Florida; Larry P v Riles; *National Consortium on Testing; New York (New York); South Carolina

ABSTRACT

Remarks made at the conference are summarized in seven sections of this narrative report: (1) update: recent events concerning testing,--representing the viewpoints of school districts, professional associations, government agencies, and research organizations; (2) National Institute of Education (NIE) study of school districts' use of test and evaluation information (Mary Ann Millsap); (3) Florida minimum competency testing case: Debra P. versus Turlington (Diana Pullin and George Madaus); (4) relationship between testing and curriculum (a panel: Joel Weiss, Decker Walker, Ann Cook, Andrew Porter, and Amity Buxton); (5) recent legislation to regulate testing (Walt Haney, Rex Brown); (6) test coaching controversy and the Federal Trade Commission report (Walt Haney); and (7) summary and discussion (Walt Haney, Eugene Radwin, and Vito Perrone). The appendices include the Speech Communication Association's criteria for evaluating tests and assessment guidelines; "The Essentials of Education," a statement from 12 organizations; a description of the Math/Science network; Banesh Hoffman's criticism of mathematics test items; a summary of the court decision, Larry P. versus Riles; a description of the NIE study and of further grants; sample materials from a school district's program; and a description of the impact of New York state's La Valle bill. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED187741

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

SUMMARY OF THE FALL 1979 CONFERENCE

OF THE

NATIONAL CONSORTIUM ON

TESTING

November 5-6, 1979

Prepared by

Walt Haney

&

Eugene Radwin

Huron Institute
123 Mount Auburn Street
Cambridge, Mass. 02138

December 1, 1979

This summary has been prepared from notes and partial tape recordings of the substantive meetings of the conference. It does not represent a verbatim account of the conference.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Paul L. Houts
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM800 237

TABLE OF CONTENTS

PAGE

I INTRODUCTION 1

Vito Perrone

David K. Cohen

II UPDATE: RECENT AND UPCOMING EVENTS CONCERNING TESTING 3

- Carolyn Del Polito, Speech Communication Association
- Dennis Holmes, Department of Defense
- John Maxwell, National Council of Teachers of English
- Lucy Sells, Math/Science Network
- Banesh Hoffmann, Professor Emeritus, Queen's College
- Bill Rosser, Chicano Education Project
- Leslie Law, Virginia Beach Public Schools
- John Weiss, Project to DE-mystify The Established Standardized Tests
- Walt Haney, Huron Institute
- Ann Kahn, National Parent Teachers Association
- Debby Meier, Community Resources Institute
- Eva Baker, Center for the Study of Evaluation, UCLA
- Francis Quinto, National Education Association
- Homer Elseroad, Education Commission of the States

III NIE'S STUDY OF SCHOOL DISTRICT USE OF TEST AND EVALUATION INFORMATION 9

- Mary Ann Millsap, NIE

IV THE FLORIDA MINIMUM COMPETENCY TESTING CASE; DEBRA P. V. TURLINGTON. 11

- Diana Pullin, Harvard Center for Law and Education
- George Madaus, Boston College

Table of Contents (cont.)

PAGE

V THE RELATIONSHIP BETWEEN TESTING AND CURRICULUM: A PANEL 19

- Joel Weiss, Ontario Institute for Studies in Education
- Decker Walker, Stanford University
- Ann Cook, Community Resources Institute
- Andrew Porter, Michigan State University
- Amity Buxton, Oakland, California Teacher Shelter

Discussion

VI REVIEW OF RECENT LEGISLATIVE PROPOSALS TO REGULATE TESTING 49

Walt Haney, Huron Institute

Rex Brown, Education Commission for the States

Discussion

VII THE TEST COACHING CONTROVERSY: WHAT THE FTC REPORTS DID AND DID NOT SAY 54

Walt Haney, Huron Institute

Discussion

VIII SUMMARY AND DISCUSSION OF SMALL GROUP SESSIONS 65

- Plans for Case Study of South Carolina Basic Skills Assessment Program, Walt Haney
- Plans for Case Study of New York Citywide Reading Test, Eugene Radwin
- Discussion of NCT Collaborative Study of Test Use and Development of Materials on Testing, Vito Perrone

ATTACHMENTS

I Speech Communication Association's "Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening"

II Statement sponsored by twelve organizations:"The Essentials of Education"

III Description of the Math/Science Network

IV Banesh Hoffmann's letter regarding flaws in a math competency test.

V Summary of federal court decision in the case of Larry P. v. Riles, regarding IQ testing.

VI Description of NIE-sponsored Study of School District Uses of Evaluation and Testing Information

VII Letter from Mary Ann Millsap describing NIE grants announcement soliciting grant applications on issues relating to uses of testing and evaluation information.

VIII Sample materials distributed by Amity Buxton.

IX Summary of New York State Department of Education Report regarding initial impact of New York's admissions testing law.

I. INTRODUCTION

The 1979 Fall meeting of the National Consortium on Testing was held at the Arlington Hyatt House, Arlington, Virginia on November 5 and 6. This report summarizes the discussions and presentations made over the two days. A separate report on the business meeting of the NCT members has been prepared and distributed.

Vito Perrone, the president of the Consortium, opened the Conference by noting a problem of success. As the NCT has grown to its present membership of 39 organizations, it has come to include a variety of perspectives on assessment and evaluation issues. Although the resulting tensions are chiefly "constructive," they sometimes may produce confusion or frustration, particularly when the NCT hasn't, as some thought it might, "changed the world over night." To allay any such concerns, Vito reminded the participants that considerable consensus exists as to the goals of the Consortium--promoting public understanding of tests, developing alternative assessment means, considering policy issues that relate to testing, and promoting research on assessment and evaluation. And he also reminded the members that they entered "the fray for the long haul."

Vito noted that the NCT was completing a two year cycle of funding from the Carnegie Corporation. The support, which has been used for staff resources, informational activities and policy related initiatives, was expected to be renewed. The new funding proposal, of which members had copies, reflected concerns expressed during prior conferences and included the undertaking of case studies on relevant testing issues. The case studies, two of which have just

been begun (and which are discussed in more detail below), Vito believes, are moving the NCT closer to one of its early goals: the examination of testing practice which would provide an additional basis for information sharing and discussion among members.

In conclusion, Vito asked the NCT members and others concerned with testing, to attend to the erosion of support for equal educational opportunity that seems to have accompanied shifts in the racial and economic composition of public schools: "One of our challenges is to assure, I think, that testing not become the means for justifying flagging commitments to public education that are already creeping into national thought."

NCT's executive director, David Cohen, briefly described the recent activities of the Consortium staff. He drew members' attention to the fact that their information packets for the Conference contained copies of the just completed sixth NCT Staff Circular, Andrew Strenio's "The Debate Over Open Versus Secure Testing: A Critical Review." David suggested that the paper -- for which Andy later thanked Consortium members for their assistance in preparing -- would likely become the "focal point' for debate on open/secure testing issues. As part of the Consortium staff's involvement with these issues, David noted that Walt Haney had testified before Congress on the Weiss Truth-in-Testing Bill. David noted that one significant Consortium activity was the preparation of the refunding proposal to the Carnegie Corporation. On a more substantive level, several projects that are currently being developed were also mentioned by David. The newest are the two NCT inspired case studies that Vito had referred to and that would be discussed in depth at later conference sessions. One project, being undertaken by Walt Haney, is focusing on South Carolina's early childhood testing

program, while the other, which is being funded by the Ford Foundation and which is being conducted by Gene Radwin, will examine New York City's Citywide Reading Achievement Testing program. Finally, David noted two informative studies by Walt Haney and Laurie Scott. One, an analysis of the most commonly used achievement tests, is close to completion (drafts of the analysis have been distributed for comment and criticism). The other gathered young children's perceptions of why they choose particular test answers in an effort to better understand issues of potential ambiguity in test items. David found the results provocative, particularly in the considerable degree of intelligibility that the items had for the children studied. The results of both examinations, Walt suggested, would be ready for distribution by the next NCT conference.

II. UPDATE: RECENT AND UPCOMING EVENTS CONCERNING TESTING.

Following tradition established at past NCT conferences, the first session of the conference was devoted to an update and sharing of information among Consortium members.

Carolyn Del Polito, described recent work by the Speech Communication Association regarding assessment of speaking and listening skills. She distributed copies of her organization's document,"Criteria for Evaluating Instruments and Procedures for Assessing Speaking and Listening," (see attachment 1).

Dennis Holmes from the Department of Defense (DOD) then described a joint NIE/Department of Defense project designed to facilitate teachers' use of test data. The project will aim at helping teachers in DOD overseas schools, but also should be of help to others. When Dennis noted

that the focus of the project would be on existing standardized tests, Judah Schwartz questioned whether the project would be more useful if broadened to include alternative types of assessment.

The National Council of Teachers of English representative, John Maxwell, reported that since the last NCT Conference 12 organizations, some of them members of the NCT and most of them sharing a subject-matter orientation had become involved in the preparation of "Essentials of Education," a statement of educational goals that is "broad in language but specific in intent" (See attachment 2). The document stipulates an unwillingness to accept popular definitions of the "basics" as sufficient to define educational goals and needs. The piece, which John characterizes as "philosophical" and "political" in nature has important implications for testing in that it touches on broad areas of behavior and learning, for which there are no assessment instruments. John invited interested individuals and organizations to become involved in distributing the document and in the Organization of the Essentials of Education, which is a consortium of the groups involved in preparing the document. He noted that OEE plans to prepare future pieces elaborating essentials in mathematics, reading, and the subject areas. Those interested can contact John through the National Council of Teachers of English.

Lucy Sells explained how the Math/Science Network (See attachment 3) is dedicated to informing students, parents, teachers, administrators and the public about the importance of mathematics and the sciences for equal education: "Without algebra and geometry, people are shut out of skilled technical careers; ... [a lack of] calculus ... closes the door on careers in engineering, business administration, computer sciences, premedicine...." Lucy also

shared her group's concern about the disproportionate impact of tests on minority-group students. This tendency raises questions, she argued, about the validity of tests, about the pervasiveness of the self-fulfilling prophesy, and about the health care provided to poor families.

Banesh Hoffmann brought to the group's attention two aspects of current testing practice that may cause students who understand a concept to incorrectly answer questions about that concept. In the first instance, Banesh cited a test item that included a "deliberate trap" among its multiple choice answers -- that is, a choice which would come up in calculating the answer but which was one step from the true answer (See attachment 4). Banesh argued that (1) such choices not be counted the same as answers which involve no understanding of the problem and that (2) children's selection of such choices may represent the effects of the stress of testing rather than a depiction of the testee's knowledge. One member of the audience observed that such examples show the importance of considering the nature of a testee's answers and not just whether they are correct or not. Banesh's other concern was exemplified by the following test question, from the National Assessment of Educational Progress: "One rabbit eats two pounds of food each week. There are fifty-two weeks in the year. How much food will 5 rabbits eat in 5 weeks?" From such questions, Banesh believes, we can infer something of children's gullibility or trust, but nothing of their mathematics ability. As in the case of deliberate distractors, irrelevant information may befuddle children into selecting an answer not reflecting their knowledge.

Bill Rosser, of the Chicano Education Project, related how his group is working with state legislators, NCT staff and members, and testing experts to develop model minimum competency test legislation. The have put together several drafts and plan to distribute a subsequent draft for comment within the next few months. In general, the Project is concerned with the impact of competency tests on minority students and, at the same time, is intrigued by the potential of "decent testing, if such is possible."

Leslie Law described a directory she has put together which specifies where and how to locate information about various aspects of testing. Leslie asked members' help in securing a publisher for the directory. Members may obtain a copy of the material for a minimal fee from Leslie at the following address:

Leslie Law
Virginia Beach Public Schools
P. O. Box 6038
Virginia Beach, Va. 23456

John Weiss of the Project to DE-mystify The Established Standardized Tests invited NCT participants to a conference on testing reform that his group was about to hold. He also invited members to subscribe to and to send material to the Testing Digest, a bulletin published regularly by the Project which has a national circulation of about 20,000. John noted that the National Public Interest Research Group has developed a comprehensive packet on truth-in-testing issues and legislation which is available for $2.50. And finally, he reported that plans are being developed for a series of teach-ins on testing at college campuses across the country for next spring. Further information about the Testing Digest, the truth-in-testing information packet, and the teach-ins can be obtained from the Project to DE-mystify The Established Standardized Tests, 1129 21st NW, Washington, DC 20026.

The recent decision in the California IQ testing case (Larry P. v. Riles) was reviewed by Walt Haney (See attachment 5). On Oct. 16, the judge ruled -- in this eight year old case -- that IQ tests are biased against minority group children and barred their use for placing school children in special education programs. According to Walt, a 500 page summary prepared by the plaintiffs (from a 10,000 page trial transcript) details the history of testing and special education in California. While the decision directly affected only Northern California, Walt predicted it would have widespread ramifications. Amity Buxton noted that, as a result of this case, the California State Education Department is already looking for alternative testing measures. Because of the interest of NCT members in the case, Walt will arrange to have copies of the decision available at a reasonable price.

Ann Kahn from the National Parent Teachers Association reminded conference participants that the PTA has recently published a booklet, "The Testing Maze," which examines standardized tests, alternatives, uses of testing, etc., and which can be obtained for $2.00 a copy from National PTA, 700 Rush St., Chicago, Il. 60611. The booklet is excerpted in the November issue of the PTA's national magazine, PTA Today. This issue, in fact, focuses on "parents testing tests." Ann sees these activities as representing the PTA's growing interest in testing issues. She noted that at recent PTA conventions, workshops on testing were well attended.

Deborah Meier directed members' attention to the needs of the Prospect Archives. The Prospect School in Bennington, Vermont has maintained a collection, now dating back 14 years, of the art work and writing, produced by its pupils as well as teacher observations, interviews, and other materials

characterizing children's abilities and development. The School is seeking support to preserve and catalogue the materials it possesses. Debby said that people who are interested in learning more about the Archives can write to Prospect School, North Bennington, Vermont, 05257. Vito Perrone mentioned that he hopes that Pat Carini, the School's director, can make a presentation at a future NCT conference.

Eva Baker, representing the Consortium's newest member, the UCLA-based Center for the Study of Evaluation, described several projects of the Center, including a long-term study of test use, research on the costs of testing, and research on the design of tests.

Frances Quinto reported that the NEA will be holding a series of 3 different workshops in 6 different locations across the country on testing issues -- while these are organized for teachers, the public is invited to attend. She also noted that the NEA, through its Program for Professional Excellence, has published a variety of materials on testing, including Teachers and Testing, Looking at Alternatives, Parents and Testing, and Ethics and Testing. (These are available from the NEA, 1201 16th St., N.W., Washington, D.C., 20036.)

The Education Commission of the States, according to its NCT representative, Homer Elseroad, has become more concerned about testing and evaluation matters, and intends to be a positive and active force in regard to testing issues. As of Oct. 1, ECS organized a Center for Assessment and Evaluation, directed by Roy Forbes. Rex Brown, who will be working with the Center, has received a Ford Foundation grant to do a study of truth-in-testing legislation, which should be completed in early 1980.

III. NIE STUDY OF SCHOOL DISTRICT USE OF TEST AND EVALUATION INFORMATION.

Mary Ann Millsap of the National Institute of Education briefly described a new NIE sponsored study of school districts' use of test and evaluation information (See attachment 6). She reported that the study, which will be carried out by the Huron Institute, consists of two related activities. First will be case studies of twelve school districts which have successfully tied evaluation and testing information to instruction and to educational management. Second will be the development of technical assistance strategies to help other school districts in better using evaluation and testing information. She noted that in the first phase, the study will focus on districts that successfully use test and evaluation information in technically sound ways. The study project is currently engaged in identification of districts which do in fact do this.

Mary Ann also briefly described the current NIE grants competition. She pointed out that the NIE's Program on Testing Assessment and Evaluation is particularly interested in funding research activities related to uses of testing and evaluation information in school districts. She further invited anyone interested in this grants announcement to contact either herself or Corinne Scott at NIE (See attachment 7).

Following Mary Ann's presentation, several questions and comments were offered from the audience. One person asked how the current study of test and evaluation use related to the second phase of the NIE-sponsored study of minimum competency testing. Mary Ann noted that the second phase of the competency testing study was up-in-the-air since a decision had been make to conduct a separate competition for it. She noted that the first phase of the competency testing study would be completed in December, and then

sometime in the spring NIE would hold an open competition for the second phase, rather than simply having the original contractor, National Evaluation Systems, carry out the second phase.

Jim Fallows commented that he's heard numerous anecdotes about successful test use, and asked Mary Ann what she thought of such accounts. She commented that while such anecdotes are available, we really do not have much good evidence regarding successful use of test and evaluation information.

Susanna Doyle inquired as to the basis on which it was assumed that there are any districts using test and evaluation information successfully. In a related comment, Judah Schwartz said that he was troubled by the apparent lack of concern for what he viewed as a zero-order question, namely what is the quality of the test and evaluation information typically available to school districts. Mary Ann commented that the study would be giving consideration to the technical quality of test and evaluation information, and said that it was precisely because of uncertainty on such issues that the study was being carried out in two phases.

Lillian Weber said that she was bothered by the implication that the study might focus mainly on technological judgments regarding tests and evaluation without giving sufficient attention to their practical consequences and their effects on teaching and learning. She suggested that it seemed as if the study might be slighting the descriptive aspect of what is really happening in schools with testing and evaluation. Mary Ann commented that this was a good point and that they would indeed need to look closely at a variety of assessment and evaluation procedures and not just

focus on standardized tests.

Ann Kahn asked whether the study would consider the framework within which school districts can develop their own assessment and evaluation approaches. She suggested that it is important to consider the barriers which exist to prevent school districts from developing their own assessment and evaluation procedures for their own purposes. Mary Ann agreed, saying that it would be important to consider such barriers, for example the effects of state and federal reporting requirements and the constraints which these place on school districts.

Banesh Hoffmann inquired as to what criteria of successful use would be used in the study. Mary Ann answered that the criteria would be related to a variety of issues related to both instruction and management of education. She commented too that she couldn't answer that question in much detail because the study was just getting under way, and because she was not really an expert in testing, but was more knowledgeable about evaluation.

IV. THE FLORIDA MINIMUM COMPETENCY TESTING CASE: DEBRA P. V. TURLINGTON

Early last summer a federal judge placed a four year moratorium on the implementation of a Florida State law requiring students to pass a minimum competency test to receive a high school diploma. The nature and meaning of this important court case and decision were discussed by two key participants at the trial: Diana Pullin of the Harvard Center for Law and Education who was one of the plantiffs' lawyers and George Madaus of Boston College who provided expert testimony for the plantiffs.

As Diana described it, Florida's Education Accountability Act of 1976

was an early manifestation of the minimum competency testing movement. Through the law, Florida's legislature added two requirements to existing public high school graduation standards: (a) proficiency in an examination of basic skills and (b) proficiency in functional literacy. The recent law suit, Debra P. v. Turlington, focused on the law's second provision. However, as more than 3,000 students, mostly from minority groups, have been denied graduation on the basis of the first provision, it is likely that this too will be challenged in court.

The law's second requirement, proficiency in functional literacy, was implemented by Florida's State Department of Education in terms of a test which, according to the state, measured the applications of basic skills to real-life situations. The test was first given in the fall of '77 to all 11th graders and then in the fall of '78 and the spring of '79 to those who failed it on prior administrations.

On the first administration, there was a failure rate of 36%; but 78% of black students failed, while only 25% of all whites did. This "significant disproportion" repeated itself on successive administrations. Diana noted that on the high school graduation day in spring '79, a black student had a ten times greater chance of being denied a diploma than a white student. It was this sort of discrepancy that led the plantiffs to file their case (474 Fed. Supplement 244) in October 1978.

Over the course of the four week trial in the spring of 1979, the lawyers for the plantiffs grounded their opposition to the use of the test to deny diplomas on two legal theories. First, and most traditionally, the disproportionate results by race on the test were used to support claims made on

the basis of unlawful racial discrimination arguments. In essence, it was maintained that (1) "the effects of the use of the tests perpetuated past racial discrimination in education. . . ;" (2) that the test contained culturally biased items; and (3) that in the way the results were used to place children, they served to resegregate public schools. Second, a set of "due process" arguments were presented to the effect that the test needed to but did not reflect or "match" the high school curriculum; neither students, parents, teachers, nor administrators had sufficient time to provide a curriculum and instruction in the areas on the test. In addition to these arguments, there was considerable testimony on the technical qualities of the test reliability and validity. Diana described it as a "battle of experts for half of the trial."

The eventual outcome was helped by the findings of a state commission appointed to investigate the new legislation. This commission noted the lack of match between the mandated testing and the curriculum. The commission actually recommended a one year moratorium on the use of the test, but this was overturned by the state education department and rejected by the legislators.

While the commission's recommendations helped the plantiff's case, the trial jduge's decision to place a four-year moratorium on testing was based on the case made on the plantiffs' behalf. The judge, accepted part of the racial discrimination claim. Children now in the upper secondary grades had been subjected to racially segregated schooling in their first few grades. The four year moratorium would assure that the test would not apply to any student who had experienced any education in

segregated schools. However, the judge ignored, in Diana's words, "massive testimony" on ongoing effects of segregation. He also rejected claims about the cultural bias of test items. In regard to the unique "due process" claims, the court affirmed the validity of awarding a diploma on the basis of achievement on a test; however, the court also affirmed the need for a match between the curriculum and such a test.

Disappointingly but not surprisingly, according to Diana, the court did not base its decision on the quality or development of the test. In a footnote to the decision, the court noted deficiencies of "considerable magnitude" in regard to the test itself, but it noted a lack of a legal theory upon which to rule on this basis. Diana later said that she sees this as reflecting the judiciary's general hesitancy to become involved in what are perceived as educational policy issues.

The trial of the Florida case is not yet over. The State Department of Education initially, and later the plantiffs, have appealed the decision. Moreover, as noted above, litigation is likely in regard to the other examination mandated for high school graduation.

George Madaus, who testified for the plantiffs on the relationship of the Florida examination to the APA test standards, reviewed some of the technical issues which arose during the trial. He noted that there seemed to have been no standardization for the administration of the test; for example, it was given over a two to three week period; also it was intended as an untimed test, but in some locales it was given as a timed instrument. There were many suggestions that the test security was compromised in some locales around this state. There were also problems in the development of the test.

Documents revealed that test items were written prior to the item specifications for one draft of the test. Altogether test development involved the Florida State Department of Education, the Dade County School System, ETS, and NES of Amherst, Massachusetts. At a different level, problems arose because the state viewed the test's reliability in terms of the traditional concept of internal consistency. However, George believes that the real reliability issue is the "accuracy of classification" -- false positives and negatives. The state made no effort to determine the level of this problem.

George's overriding concerns both at the trial and in his discussion at the conference were with the issues, which interweave, of content validity and inference. George views the test as a certification instrument, and consequently believes the test items should relate to instruction and curriculum. For him, content validity in such a context means that test items must not only accurately represent their domains but also that the item domain must reflect the curriculum. The state's experts disagreed, denying the necessity of the domain/curriculum linkage. George related how in response to a hypothetical situation proposed by Diana, one of the state's witnesses held that if Florida required mastery of Greek for graduation, even though it didn't provide for instruction in Greek, test items would have content validity for certification as long as they represented the domain of the Greek language.

George believes that the Florida State Department had to defend such a tenuous proposition because of the way the legislation is written; the test is mandated to be a test of functional literacy. Although the state has

dubbed it the SSA II test, it is known popularly and on official documents as a functional literacy test. The test carries the weight of literacy on it; children who fail are thought to be functionally illiterate; and in spite of official denials as to their existence, the state conducted several studies based on the presumption that the instrument assessed literacy skills. On one such trial, George noted parenthetically, 30% of the adults who took it failed; these were all employees of the state education department. For George, the construct of literacy/illiteracy is clearly involved in the test, and therefore, if such inferences are being made, it is encumbent on the state to establish the validity of them; however, the state denied this logic, claiming that the test just tested the application of basic skills to everyday problems. The State denied specifically that it was employing the functional literacy construct.

In spite of all these problems, George stated that he came away from the trial with ambivalent feelings. "Deep down, with all the faults of the test, every 11th grader in the world should have been able to pass it . . . . But to me it was a terrible indictment of the educational system that kids at the 11th grade couldn't answer 70% of these items. What Florida chose to do was very interesting. They decided to put the onus on the kid and hold the kid's diploma for ransom in hopes that this kind of pressure from an external test would focus instruction, and that if we were willing to sacrafice a few cohorts along the way, eventually kids would pass these kinds of tests. It is a clear example of using tests in public policy as an administrative device to get something done that you couldn't get done otherwise; it's taking control away from local school districts and putting a very powerful mechanism in the hands of the state department; and those people I

think are people of good will; they see that this is one way to get curriculum reform. A very tough way on those kids though."

In a final comment, George said that he was struck by the lack of organized parental lobbying, especially among minority group parents, opposing the test. He wondered if they understood the test and the issues it raised. To deal with this testing program he suggested, we have to address it as a political issue but local groups didn't seem prepared for this perspective.

The question period following George and Diana's presentation was a lively one. Paul Jacobs asked how the passing grade on the test was established. Diana responded that a 70% level was set after a limited phone survey of principals and superintendents found that this level was the most commonly used one and was one parents and students could understand. After the fact, there was an attempt to use more sophisticated procedures to justify this level.

In response to Banesh Hoffmann's questions about the arguments the judge discussed in his decision and the judge's reaction to disagreements among experts, George and Diana made several points. For one thing, the judge worked hard to understand the issues. More specifically, he accepted functional literacy as a concept but left it to the state to define in any which way. In regard to reliability, he accepted the fallibility of the test or any other test and was only concerned with the number of false negatives. Finally, the judge was caught between the experts. He apparently could not rule against the test on technical grounds -- as much as he seemed to agree with the plantiffs -- as long as the state offered any testimony defending the instrument.

Laura Hines of the APA pointed out the kinds of obstacles that may have interfered with parents organizing an anti-test lobby, including lack of money, knowledge, and time. She stressed that putting together such a lobby is a difficult task. Diana agreed and admitted she was frustrated by lacking the time and resources to disseminate information about the test. She noted that the trial did have some impact, and because of it public sentiment was no longer so fully supportive of the testing as it had been.

George discussed the curricular impact of the Florida legislation in response to a question on this matter from David Cohen. Remediation programs have been set up to help children who are at risk or who have failed the test, but complaints have arisen that these are pull-out programs and take time away from science and higher mathematics instruction. Similarly, complaints have been registered that the emphasis on the basics is having an adverse affect on music and arts instruction. There are also reports, according to George, of teachers being held accountable on the basis of their students' performance on the test. In addition, Florida is spending considerable money to develop instructional modules geared to the test. However, these have yet to be evaluated as to their effectiveness or to their extent of use. Along these lines, Diana noted that near the end of the trial, the Florida legislature passed a law allowing kids who failed the test to return to school for a 13th year. The students, however, felt they should have received the necessary instruction in the 12 years designated for a public education.

Another questioner, stressing the grass roots nature of minimum competency testing, asked the presentors about the impact of the decision, about

whether the decision was worth it, and what the future held? George stressed that the situation was predictable: if you have an external test with severe sanctions on the pupil, teacher, or school, the agency that sets the test has considerable power; eventually what the test emphasizes will be incorporated into the curriculum and test scores will rise. The model is a European one except that it is not fully meritocratic; rather it is a minimum one. Diana stressed that the problem was that the test wasn't the creation of a grass roots movement. As it was, a small number of education department employees worked out what the test would be. There was a lack of public visibility.

The final question of the session was posed by Marjorie Martus of the Ford Foundation who asked about the relation of the Florida case to the "opportunity to learn" issue that has been raised in several court cases. Diana noted that educational malpractice cases have largely failed because courts have lacked standards for defining what appropriate criteria of educational practices are. However, the minimum competency testing movement may be moving education in that direction. The tests define a level of achievement for schooling that could in time be applied in court cases. Thus, she commented through minimum competency tests education may be "hanging itself" in terms of future litigation.

V. THE RELATIONSHIP BETWEEN TESTING AND CURRICULUM: A PANEL

After noting that the issue of the match between testing and curriculum had already arisen in discussions of minimum competency testing, Vito Perrone introduced the panel of individuals invited to address the various relationships between testing and curriculum. The first panelist was Joel Weiss of

the Ontario Institute for Studies in Education.

The task of analyzing the relationships between testing and curriculum, Joel suggested, was a formidable task. He recounted that at least three perspectives could be helpful in undertaking such an analysis: 1) the curriculum, what ought to be done; 2) the technical, what can be done; 3) the political, what will be done. Any analysis of testing should take into account these three facets of the issue.

Joel suggested that the Follow Through (FT) program and evaluation represented a good example of the interaction of these three facets of testing. First, FT encompassed a range of curricular approaches to early childhood education. Second, the FT evaluation was driven, often at the expense of program considerations. by technical considerations. Third was the stress on the widely accepted dichotomy between cognitive and noncognitive outcomes, which might not make much sense from a curriculum point of view. The instruments used in the FT evaluation, he suggested, were really based on convenience. In short the evaluation seemed to be based on a sort of least common denominator. Moreover, he suggested the ensuing political controversy regarding the FT evaluation shows the political vulnerability of the relationship between curriculum and testing.

Also, he suggested, another important perspective on the relationship between curriculum and testing is opportunity to learn. Consideration of opportunity to learn helps to make clear the contrast between rhetoric and real curricular experience. Although we often talk of children being exposed to a whole range of curriculum topics, in practice they may not,

Thus, we have two kinds of mismatch between testing and curriculum -- the appraisal of non-events, and the non appraisal of actual events. Both problems were represented in the FT experience, Joel commented, and also may be represented in treatments of the National Assessment. Results of National Assessment are often compared over time without consideration being given to children's opportunity to learn in various areas. Assessing opportunity to learn, he concluded may be just as important to assess as test outcomes.

Another point Joel made was that although everyone may have their own curricular biases, focussing on a relatively narrow range of test outcomes may be quite common. In the research on open education, for example, two-thirds of the studies recently reviewed dealt primarily with achievement, self-concept and attitudes toward school. These seem to be the three outcomes, which have come to pass for current educational value. Even so, most of the studies focused on achievement of the sort covered on standardized tests, with relatively little attention given to the sort of things open educators are most interested in doing with children. This contrast, he suggested, derived from the perceived political necessity for such programs to prove themselves on the grounds of traditional schooling, to show that children in open education can read, write and compute just as well as children in traditional schools.

Joel continued by noting that when we turn to the literature from the progressive period -- bounded roughly by the 1920s and 1950 -- a different

pattern emerges. First, not nearly as many studies were done. Second, studies were in general far more comprehensive. Their vision of what we expect of children in school was far greater. So these studies appear to have been far more in tune with the curriculum philosophy of the progressive movement.

What are the implications of this contrast, Joel asked? What kind of study would Neville Benett have done if he were more philosophically in tune with the practices of the teachers he was studying? Clearly, there are real differences among the researchers, evaluators and testers. In the progressive era, investigators had a real sense of scientific advancement and really believed in achievement testing without getting a real sense of what was going on in the schools, what teachers were trying to foster under the name of progressive education. They were aware of the political necessity of selling progressive education, but at the same time they seemed to truly believe that scientific experimentation would prove the benefits of progressive education. In contrast, Joel remarked, this does not appear to be the stance of many of the people doing research, particularly the survey research, on open education.

So there are differences among evaluators and testers. Some don't believe we should even look at the outcomes of open education, or that at least it's not as important as other questions. In some ways, he suggested, testing of achievement seems to be a sort of political touchstone for showing the worth of educational endeavors.

In conclusion, Joel said, he wanted to simply raise several questions, which he hoped others might take up as challenges to be addressed. First, how can the test situation itself be used to test the effectiveness of curriculum? He described one case where students from a curriculum which had attempted to make them more aware of subject matter distinctions and had stressed honesty and self-examination completely refused to take the semantic differential instrument. They refused because they felt a semantic differential was completely out of phase with the subject matter they had been exposed to, and was an indirect assessment of attitudes, feelings and values which ought to be addressed directly. Since the kind of testing we do may sometimes be out of phase with the kind of things we're trying to test, Joel went on, children's reactions to testing may sometimes be important to address in our evaluative procedures.

Recounting an example from Henry Murray's book The Assessment of Men, Joel noted that there clearly are ethical problems in setting up an assessment task which assesses people on things of which they are not aware. Nevertheless, if testing is an important aspect of our curricular situation, we might usefully try to bring it more in line with other aspects of curriculum, rather than trying to suppress such differences.

As a second concluding point Joel returned to the question of the orientation of the tester, researcher, or evaluator toward the program he

or she is studying. Should such people really have a feeling for the orientation of the program they are studying or can we get around that by assuming that regardless of what program they're considering there are a whole range of outcomes to look for? Clearly this presents a dilemma, Joel suggested. But there do seem to be instances in which detrimental consequences ensued from the fact that investigators lacked a curricular sense of programs studied.

The third concluding comment by Joel dealt with what actually happens with policy research. He suggested that we may be fooling ourselves by including nonconventional outcomes in large-scale evaluation and pretending people pay attention to them. He suggested that instead of following this approach we might better reserve resources for addressing non-conventional outcomes on a case study or formative basis. At the same time we might move away from the pretense that by including such nontraditional outcomes in policy-oriented studies that it's going to make a difference.

The second speaker was Decker Walker of Stanford University. The questions he addressed were whether testing is driving the curriculum, and the reverse of that question, whether curriculum is driving testing. Decker spoke, he said, as a student of American curricular history, structure and problems, not as a student of testing. Tests do, he suggested, affect what children do, and what they learn and tests also affect the curriculum They do so, however, in complicated ways which we do not completely understand. We know that the effect of a test does not simply result from the test as a piece of paper, but that the effect is conditioned by such things as: who made the test (the kid's own teacher or some multinational corporation); and where the child heard about the test (from peers, family, or only from the teacher just before taking it). We know its effect depends

on test format, as Banesh Hoffman points out. We know it's affected by the conditions under which the test is administered. We know it's affected by the consequences that are made conditional upon results of the test. We know also that it is affected by the perceived consequences of the test. We know that it is affected by internalized meanings which youngsters may place on the test. So the situation with respect to tests, and their direct effect upon the youngster's allocation of time, energy and attention, is very complicated.

The situation is at least as complicated in the case of the curriculum's effect on the youngster's time, energy and attention. And in addition there are other effects of the testing on the curriculum, and the curriculum on the testing, and other factors and influences that affect both.

Decker then proceeded to elaborate some of those complications. His general theme was that it's a complicated situation, complicated for good reasons and it's not likely ever to become simple. As a result there are no simple solutions to the problems that arise. Nevertheless, he suggested some things which could be done to help move us forward and enable us to resolve some of the worst problems.

First, Decker addressed the question of testing's effect on the curriculum. Reading achievement tests, for example, have led to the establishment of grade level norms for performances on these tests. These in turn have led publishers of textbooks to key their own text material to the reading level of these norms. So if youngsters in the eighth grade are reading at the sixth grade level, publishers find themselves under great pressure to key their texts to the sixth grade level. Thus the fact that we have such tests and norms affects the development of texts.

Another example is the fact that we have tests for entry into college

and various jobs leads to decreased reliance on high school course records, grades, and teacher recommendations on the part of colleges and employers. As a result, among many others, the influence of colleges and universities on high school curricula is reduced. Whereas colleges previously required candidates to have a familiarity with certain ideas or works of literature, they now can simply require a certain score on the SAT. This leads to a certain independence on the part of students from individual schools' grading or evaluation procedure, because the student can always rely on getting a certain score on the SAT and thereby getting in to college. Whether one thinks this is good or bad likely depends on his or her point of view.

Tests, being paper and pencil matters, also are not that usable in some subject areas -- the arts and physical education for example. The fact that they are not and the fact that there are tests used in other subjects areas in very important ways for influencing life decisions tend Decker suggested, to devalue the place of those less easily testable subjects in the school curriculum. There is simply no way for accounting for them in accountability schemes based on tests.

Decker went on to illustrate a larger loop in the complex of connections between tests and curriculum -- the use of tests in research designed to influence public policy. The classic example of this, he said,is the Coleman study which affected policy on American education for a decade and echoes of which are still being heard. The tests which provided the data for conclusions about which there was most debate were adapted from the verbal portion of a standardized IQ test. The fact that the test was used in that way as an indicator of how well

American schools were doing is one manner in which tests influence curriculum, through influencing policy discussions and policy debates. In a just world, Decker said, he doesn't think much attention would have been given to the Coleman report for the simple reason that it was based on such flimsy raw data -- and tests not useful for the purposes to which they were put.

Moving from illustrations of the tests' effects on curriculum Decker noted that curriculum also affects testing. Some ways in which this happens are obvious -- for example the keying of items on tests to the content of current textbooks. Another example comes from the introduction of new science curricula in the post-Sputnik era. These new curricula covered things not on the tests, so the developers of the new curricula faced quite a chore in trying to persuade the testers to include items reflecting such curricula on their tests. Decker recalled his personal outrage, as a then science teacher, that testers should be so conservative and not change. He also recalled his reaction a few years later when new math and new science were roundly being kicked out all over the place, wondering whether the testers might not have been right in the first place.

Tests are also affected by curriculum in terms of norms, Decker continued. Items are chosen in light of item tryout data and ones that don't behave in the desired way are thrown out. The tryout is done on youngsters in our public schools. If an item is one that is universally taught and 99% of youngsters get it correct then it won't discriminate in the desired way and so will be thrown out. So norming practices are one way in which curriculum may affect the tests. Obviously developments in curricula lead to pressures on testers to change their tests in various ways.

Decker then turned to forces in society which affect both testing and curricula -- which are he suggested the largest and most important set of influences. Some such influences are obvious. The women's movement, for example, and the movement for minority rights have led to demands for changes both in tests and in curricula. Yet such forces act on testing and curriculum through two systems which are largely separate. Parenthetically he added that he probably should not use the word system to describe the means by which tests and curriculum develop, since they are not really systematic. Rather they are melanges of factors, welters of influences which lead to and mediate developments in testing and curriculum. Yet since external forces operate separately on testing and curriculum, this may lead to gaps and discontinuities between testing and curriculum.

On a more subtle, but in the long run perhaps more important scale, Decker continued, there are pressures toward standardization and formalization in all of life, but particularly in the schools, which affect both testing and curriculum. One example is the chain of prerequisites which students in high school, colleges, etc., need to meet in order to enter certain technical professions. One important cause of this trend is simply the massive scale of educational enterprises. The scale of these enterprises causes us to look for ways to simplify their operation. Imagine, for example, Decker asked, what it would be like to have to interview every incoming freshman in every subject area at any major public university, with say 12,000 incoming students. Clearly it would be unwieldy to operate in this way on such a scale. And since we insist on institutions

of these sizes we have to find simple ways to deal with them. This is, he suggested probably the primary reason we place so much emphasis on standardized tests in American education; we have to have ways of dealing with large numbers of people in relatively brief periods of time, and on the basis of relatively few bits of information. This trend toward standardization and formalization is pervasive in our society, he suggested, and it affects both testing and curriculum.

In short there are numerous effects, testing on curriculum, curriculum on testing, and other influences on both. These effects are most often noted when mismatches result between testing and curriculum. Where developments are not mismatched we hardly even notice that there is any influencing going on. There are, Decker continued, several classes or categories of mismatches which are important to distinguish. One case is where mismatches result from local variations in curricula, even though standardized testing over the last 50 years or so has tended systematically to reduce local variation. A second area is in the ratio of skill to content matter in curricula. The single most important trend since around 1890 in elementary school curriculum, Decker observed, has been the increasing emphasis on skills -- reading, writing and arithmetic. These skill subjects are, we tend to believe, the ones which are easiest to test. The content areas seem to present special difficulties for testing. What, for example, would we think of a statement that a student scored at the ninth grade level in social studies? Surely we would want to know whether this referred to history, geography, getting along with neighbors or what. Yet similar statements are readily accepted with respect to reading or math. In short there has been an increasing emphasis

on skill matter in schools at the expense of content, largely because of the better match that we have between tests and curriculum in the case of skills.

The third set of cases where there is mismatch is in cases of rapid curriculum change. Any time there is curriculum reform, testing tends to lag behind, and then by the time testing has changed, the reform has passed and people find themselves with an inappropriate test. Such mismatches may be temporary but they are nevertheless serious, especially for test-makers who have to plan on their tests being valid for five to ten years in order to get the economies of scale they need.

Testing, Decker noted, is the policy-maker's tool par excellence. It is a tool by which it is possible for one person or several in a central location to make concrete decisions that will directly affect thousands of teachers and students at a distance. In the United States this cannot be done via textbooks, as it is in other countries. But it is possible in the U.S., at least in principle for policy-makers, including legislators or others, to exert such control through tests. This, he continued is one of the main reasons why testing has been so attractive to legislators in the U.S. in recent years. In response to legitimate concerns about educational policy, testing provides a means for them to reach down into every school and classroom, to every student and to influence them concretely and directly. So tests provide a means by which the long arm of the policy-maker can reach out and affect every student and teacher. The desire is a natural one, and certainly it is in the interests of us all to have some means of such regulation, he continued, but tests are not

necessarily an appropriate one.

So what can be done? The chief danger, Decker offered, is that of oversimplification. The danger is covering over the complexity, in which students differ from one another, subjects differ and schools differ, covering this over with one set of numbers or three sets of numbers. Drawing an analogy with ecosystems, Decker suggested that the chief danger is intervention based on an oversimplified view of the complexity with the result that the system becomes unstable or unviable.

How do we proceed then? Decker suggested a five point plan for sensibly dealing with such matters. What we need, he suggested, is discussion, connected to decisions which occur at several local levels, and at each of these levels there is some area of autonomy, and some area in which those who make decisions are required either to comply with other authorities or else to appeal their decisions. In short he concluded, what we need is a policy-making system to sort out issues of testing and curriculum. His suggestion, in short, is that the only way to respond to the complexity in testing and curriculum, is with corresponding complexity and flexibility in a policy-making system, rather than relying simply on a melange of influences. Such a system would be expensive in time and money, he noted, but if we really care about our youngsters' education we ought to be willing to put up with such expense. He noted that such a system would include much discussion among parents, students and teachers at the local level. In this respect, he recounted that as a parent he has continually had to push and pry in order to get teachers to discuss his children's performance in any terms other than grades.

In response to a question from the audience, Decker commented that

while the complexity of policymaking has been on the increase, it has been influenced in thousands of ways that most people hardly know exist. What is needed, he suggested, is more systematic policymaking so that we can make better sense of policymaking.

In response to another question from Les May concerning how policymakers at the state level could, under his system, get the information they need, Decker suggested that there is no simple answer but that what would be needed is a flexible policy-making process toward information-gathering so that issues could be considered at various levels. Thus information in detail appropriate to issues at different levels could be derived through a process of discussion and negotiation, but not through any single answer in principle nor through any sort of policy breakthrough.

The third speaker on the panel was Ann Cook of the Community Resources Institute. She discussed the impact standardized norm-referenced tests in reading have on the curriculum in reading and more generally on all subjects. She argued that the notion of reading promulgated in the schools is "indistinguishable" from that emphasized in reading tests. As one example, she noted parallel structures and questions on a CTBS item and material in the SRA Reading Labs. For Ann, this was just one instance of a more general trend: the adoption in reading series of the format and questions (short paragraphs and questions on main idea, vocabulary in context, direct information, and inference) found on every standardized reading tests. Children, Ann maintains, end up getting much practice in answering such questions but not in thinking about or understanding the reasons for right or wrong answers.

Ann questioned the impact of such materials on children's views of reading, education, curiosity, and respect for knowledge. The emphasis

on test taking skills (both in the kinds of questions posed and in the format of presentation) would seem to make potentially interesting material boring; "it is bound to develop in children an aversion not only to reading but to knowledge itself." Along this line, Ann noted how in a McGraw-Mathers workbook, interesting stories were followed by irrelevant, trivial questions.

Some other reading materials, Ann pointed out, encourage children to answer questions without reading the stories. Such practices, she believes, lead to a lack of respect for content. This lack of respect is also fostered by current remedial reading strategies. With those needing help in reading, it is most common to attempt to improve their skills by having them practice "on short, isolated paragraphs unrelated to others, taken from different disciplines and time periods, without any attempt to create continuity or sustain interest."

The kinds of materials and practices that Ann views as strongly influencing the reading curriculum, she also believes, are having a more pervasive impact. Ann cited materials from literature and social studies texts in support of the contention that the reading test format is having an impact in shaping the traditional content area curriculum. Also relevant is that with the attention given to improving reading, "we have taken all the meaning out of education; it's no longer fashionable to teach literature, history, science, logic, civics, music, or art."

One clear instance of the loss of time for content area instruction can be seen in the response of New York City schools to the annual citywide reading achievement test. According to Ann, from January onward the schools

begin to gear up for this test, which is not given until March. To provide children with practice materials, "in almost every school thousands of hours are devoted to culling selected passages or vocabulary lists from past years' tests and from commercially prepared materials." The personnel and financial resources spent by local districts are considerable. In one school, the preparation involves 25,000 mimeographed practice sheets, one secretary, four aides, and an administrator to copy, type, and collate.

A related aspect of the New York testing program, is how the curriculum in reading reflects the specific test that is being used. Thus, according to Ann, during the years that the New York City test included items on opposites, this skill was emphasized in the schools. When a new test was used that didn't have such items, the skill was dropped from the curriculum. Similarly, now that New York State has implemented tests that employ a cloze design, practice on cloze tests is included in the reading curriculum.

If we want children to become proficient test takers, Ann stated, then we should teach them strategies to translate and approach tests. Such instruction can "dramatically" improve test results. However, this has little to do with improving reading.

Ann concluded that the emphasis that is being neglected by testing, she and others see as critical; that is, "knowing something in depth; taking intellectual responsibility and defending a position; reading a whole, entire, book; critically analyzing material; and developing written and oral arguments." But as long as tests are stressed, testing will be the focus of the curriculum.

The fourth speaker on the panel was Andy Porter who presented a paper, prepared with Teri Kuhs and Don Freeman, on two NIE supported research projects that he has been directing at the Institute for Research on Instruction at Michigan State University. The first project describes teachers' willingness to alter the content of instruction as the result of external pressures including standardized tests. The second compares the content of textbooks to the content of widely used, commercially available achievement tests. In this latter project, the content of textbooks is taken as a somewhat crude proxy for the content of the school curriculum.

This research is part of a larger, systematic investigation of the distribution of mathematics content in elementary school. Throughout the research, the dependent variable is the content delivered to individual children in the classroom, and the independent variables are the factors which determine variation in that content. The eventual aim is to estimate the strength of three influences on teacher decision-making: "(1) the formal hierarchy which transmits the policy decisions of higher authorities and which, therefore, reflects political processes bearing on a school system as a whole; (2) other influences from inside or outside the agency which are brought directly to bear on the teacher; and (3) the teacher's own conceptions of what outputs are desirable and feasible." Since mathematics is less controversial than such subjects as sex education, social studies or even science, Andy and his colleagues believe that this subject will provide a good conservative test of the extent to which consideration of others' preferences and demands enter into teachers' decisions about content.

Identifying the factors that can influence content selection depends on one's view of teachers. In one perspective, characteristic of curriculum reform movements, teachers are seen as autonomous decision-makers who must be persuaded to teach certain content. In a contrasting view, consistent with the recent

accountability movement, teachers are, to quote Corwin, "employees owing obedience." In fact, these accountability and dissemination perspectives of teacher autonomy are inextricably linked; both are considered in the working model which underlies the Institute's project.

Factors external to the classroom can affect the content of instruction both directly and indirectly. Those who believe that the curriculum has deviated too far from the basics, for example, may launch a drive which results in the adoption of basic objectives for mathematics. The state may try to ensure these objectives through a criterion referenced assessment program. The direct effect may be coverage of classroom objectives. The indirect effect may be that certain other topics are dropped -- dropped because the total amount of instruction is constrained. Topics might also be dropped because they fail to gain the show of external support given the basic objectives.

The study's focus on instructional content is based on several beliefs. First, there are certain subjects which most students learn only in school. Second, it is often contended -- and the project hopes to provide some hard data on this -- that even within a subject-matter area and grade level, students vary with respect to the content to which they are exposed. Third, in the past research on teaching has concentrated on strategy and neglected content. Also teacher education has given far more attention to delivery and classroom management skills than to selecting the content appropriate for a particular student, an imbalance that may create teachers who are more interested in how something should be taught than in what should be taught.

The fourth-grade level and the mathematics content were selected as the initial contexts for investigation for several reasons. First mathematics is generally considered a basic skill that all elementary school children

should study. Also, it meets the criteria of being a subject that is primarily learned in school. In addition mathematics is a subject for which relatively precise and careful descriptions of content variation are possible. Finally, the focus on fourth grade allows for substantial variance in what might be taught.

From this general introduction to the research being undertaken at the Institute, Andy moved to descriptions of two specific projects. The study of teacher decisions about content involved 66 mostly experienced teachers from urban, suburban, and rural school systems around Michigan.

The study employed a standard "policy capturing" methodology. Subjects are presented with a series of stimulus situations -- or vignettes -- and asked to make a decision about each. The stimulus situations are each scored on one or more variables. These variables are used as independent variables in regression analyses to predict the subject's judgment. To the extent that the prediction is accurate, the resulting regression equations are taken to represent, or "capture," the "policy" that the subject used in arriving at the judgments.

For this study, the vignettes presented were descriptions of elementary schools in which the teachers were asked to imagine that they just had found a job. Their new school was much like their old school, except that most of their new students were assumed to be capable of doing fourth-grade mathematics. Each vignette described a certain combination of pressures which were always consistent in their goal, to alter the content of mathematics instruction, by having teachers add new topics or omit topics that the teacher had ordinarily covered. Six external pressures, any of

which could be either present or absent in any vignette, were of two sorts. Three of them were impersonal: testbooks supplied to the class, publication by the central administration of objectives which all teachers have been directed to follow, and publication in the local newspaper of standardized test results, by building and grade level. The other three pressures were personal: discussions with the principal, with the fifth- and sixth-grade teachers, and with parents.

After reading a vignette, subjects were asked four questions: how likely they would be to add or omit two types of topics; those that they usually taught or those that were not usually taught in the elementary school. Two of the questions asked how likely the teachers felt they would be to add content, and the two other questions with how likely they felt they would be to continue the content that hadn't been supported.

The results, for which Andy displayed several tables, suggested that the teachers believed they would add topics to the curriculum even under a single pressure, particularly if that pressure was a set of official objectives or the publication of test results. However, the teachers were reluctant (or so they conjectured) to omit topics including peripheral ones, from their curriculum. Even if one is "so cynical as to consider expressed prediction as unrelated to future actions . . . , these teacher projections can be thought of as . . . behaviors the teachers believe to be socially acceptable behavior -- the degree to which teachers see themselves going along with whatever pressures" are present in striking. Given current emphases on the

professionalism of teachers and their authority in the classroom, this is surprising whether it means that teachers are actually so malleable, or that teachers think that malleability is the socially acceptable image. In this regard tests may drive the curriculum but so do many other things.

The second study compared the content of several well known mathematics tests with fourth grade mathematics curriculum. For this work, the contents of popular fourth-grade mathematics texts were taken as a proxy for curriculum in this domain and level. This has some dangers, Andy noted. A recent nation-wide survey found that while 75% of elementary school math teachers used a single textbook predominently, only half followed it closely. Regardless, however, the match between text and test is of interest when interpreting test results.

Ideally, the knowledge domain represented by a standardized test should be consistent with the knowledge domain in which students receive instruction. Since the test represents only a sample of the domain, many more topics should be included in the textbook than are actually tested. However, as most standardized tests are designed to be used by students at different grade levels, the tests may include topics not covered by a student's textbook. It is also possible that standardized tests are designed to test for transfer of knowledge. But this is not one of the purposes of testing addressed in test manuals.

Some have claimed a national curriculum in elementary school mathematics exists, but only at a fairly high level of generality. While all students are instructed in addition, subtraction, multiplication, division, and measurement, more fine-grained analyses of textbooks, have revealed rather substantial diversity. A content analysis of three fourth-grade mathematics textbooks found that of a total of 293 specific topics, only 24 percent were covered in all three books. If one adds the requirement that the topic be represented by at least one instructional lesson, the percent of topics in common dropped to 7 percent. These data open to serious question the notion of a national curriculum in elementary school mathematics.

To guide the comparison of tests and texts, a taxonomy was developed for describing in detail fourth-grade mathematics along three dimensions; the intent (e.g.,conceptual understanding or applications); the nature of the material (e.g., whole numbers, fractions, number sentences); and the operations students must perform (e.g., add with carrying, estimate). Using this scheme, the study analyzed the following common materials: Addison-Wesley's Mathematics in Our World (1978); Houghton-Mifflin's Mathematics (1978); Scott-Foreman's Mathematics Around Us (1978); Metropolitan Achievement Test; Elementary Level, 3.5 - 4.9 [MAT]; Stanford Achievement Test, Intermediate Level I, 4.5 - 5.4 [SAT]; Comprehensive Test of Basic Skills, Level I, 2.5 - 4.9 [CTBS I]; Comprehensive Test of Basic Skills, Level II, 4.5 -6.9, [CTBS II]; Iowa Test of Basic Skills, Level 10, Grade 4 [Iowa].

Andy then discussed the data and results that followed from plotting these materials against the taxonomy. The first,demonstrated by overlaying transparencies of a taxonomic analysis of the Stanford with one of the Addison-

Wesley text, showed the areas of match and mismatch between the content test and text. From the way the transparencies were filled in, one could identify topics that were in the tests but not in the texts (such as concepts involving number sentences) and cases of the reverse (which was true for the multiplication of single digits).

The data from these overlays had been summarized by the project in several ways. One way, which is presented in the table below, consisted of simply calculating the percentage of topics common to the individual tests and texts. Andy pointed out that the SAT -- the most commonly used test in the country -- had the poorest match of any test with the texts, a finding which belies the idea of the test driving the curriculum. It was also noted that the Scott-Foresman text provided the best match with all of the tests. Such overall results, as well as those based upon considering the match at the level of subdomains (e.g., concepts), suggested that the selection of tests and texts does make a difference.

PERCENT OF TESTED TOPICS COVERED IN THE TEXTBOOK

	H.M.	S.F.	A.W.
MAT	60	71	60
SAT	51	52	47
IOWA	63	66	53
CTBS II	58	66	61
CTBS I	57	57	54

In conclusion, Andy suggested that many factors apparently help to focus the curriculum. Tests, at least on the basis of the policy capturing study, appear to be one of the more powerful. However, in terms of the test - text match, there is no national curriculum in elementary school mathematics. Even if tests are driving the curriculum, they are not driving fourth grade mathematics texts.

But even a high match between tested and taught content would not be sufficient evidence to conclude that tests drove the curriculum. To this end Andy and his colleagues are using teacher interviews and case studies of selected school districts to examine the attempts of school districts and schools to standardize the school curriculum. The researchers are investigating the methods -- such as management by objectives, curriculum experts, mandated textbooks, etc. -- which can be used to standardize the curriculum.

The results thus far indicate that these mechanisms do not share a common content message even within a school district. In the Lansing, Michigan schools, for example, the management by objectives don't match the Stanford, which is given annually, and neither match state mandated objectives.

The effect, Andy argued, is to pit one attempt for control against another in vying for the teacher's attention. The net effect appears to be that considerable teacher autonomy is still allowed. In effect, an elementary school teacher is in the position of being a political broker whose task is to decide which if any content messages are salient to classroom level activities.

The final member of the panel, Amity Buxton of the Oakland, California Teacher Shelter, described two recent efforts of the Oakland public schools to develop nonstandardized test assessment procedures. In the first instance, the Oakland system, which consists of 92 schools (62 at the elementary level) and in which minority group youngsters constitute a majority, was mandated by the state last fall to administer the CTBS (California Test of Basic Skills) to kindergarteners to identify Title I-eligible children. After teachers and psychologists saw how devastated the children were by the test taking -- many "cried" and "wet their pants" -- an effort was made, in which Amity was invited to participate, to develop another means of assessing the skills of the children.

As a result, from last November until this one, teachers in over three quarters of the kindergarten schools have been developing checklists of observable behaviors in such areas as prereading, language, auditory and visual perception, self-image, and psychomotor coordination. The teachers "developed the assessment of what they felt was the best curriculum for children; and the assessment form itself reminded teachers of all of the things they were supposed to be teaching." Moreover, the checklist helped teachers talk to parents, for it provided a focus on a broad range of specific skills. Most importantly, the state has sanctioned the use of the checklists as a substitute for the CTBS for identifying Title I-eligible children.

The other situation Amity described involved a state mandated assessment of elementary school children. State Assembly Law 65 required each local district in California to "detect" early on -- between grades 4 and 6 -- children's proficiencies in those areas required for high school graduation.

The proficiency areas and assessment instrument were left to the local districts to specify.

To meet the requirements of the law, Oakland decided to draw upon the standard embodied in what are called "grade-level learning expectations." These expectations, which were calculated on a grade-by-grade basis, were devised several years ago by teachers and principals and were designed so that most of the children could get most of them right. After the passage of Bill 65, the district decided to assess the skills of the 3rd and 5th graders; the district also decided that proficiency on "key" grade-level learning expectations would be required for promotion from these grades. Initially, the district intended to employ either the CTBS or a locally developed criterion referenced test to assess such proficiency; however, neither instrument was found to adequately cover the necessary material. At this point, Amity was invited to help, and she and a group of teachers spent last summer devising an alternative assessment process.

The group decided that the assessment procedure would incorporate all the designated key expectations, would subsume all proficiency standards within these key areas, and would build the entire process into the curriculum. (A sample set of the materials developed for the resulting assessment procedure is included in attachment 8.) In essence, Amity's group developed systematic procedures to assess children's performance on everyday curricular activities as these activities were occurring. At anytime during the year, teachers could see, according to specified criteria, whether a child's performance on an activity met performance standards. Thus, to use an example that Amity related in the

question period, if a teacher wanted to assess a child's ability to give an oral report, the teacher would wait until there was some natural class activity in which the child gave the report and then would evaluate it against specified criteria in regard to its organization, introductory and concluding statements, use of standard language, clarity, etc.

The important point, Amity stressed, was that process allowed for assessing actual behaviors in a relatively natural manner. Moreover, in that teachers were being encouraged to look at performance on more than a paper and pencil test, the process counteracted some of the curriculum narrowing tendencies that may accompany standardized testing programs. In fact, in that teachers were being asked to view children's skills across a range of activities, the Oakland assessment may be affording teachers a broad view of the curriculum and also of his or her students. The procedure places teachers' judgments above test scores in the decision-making process, but by standardizing the basis for those judgments, the procedure has met with state approval.

A brief period for questions and comments followed Amity's presentation.

Commenting on Ann Cook's presentation, Steve Ivens of the College Board suggested that assessment procedures which focus on "observable student performance" do not overcome the intrinsic problem with tests. To illustrate this point, he made the analogy that students who already appreciate art presumably can match up ten paintings with a list of the artists who drew them; however, this does not mean that students who can do this task necessarily appreciate art. For Steve, the "crux" of the problem is in the translation from a general educational goal -- such as art appreciation or

reading comprehension -- to a specific objective -- such as matching paintings and artists or selecting the main idea for a passage. Even with criterion or domain referenced tests, one can not be sure that students who perform the specific task have attained the more general goal. In fact, once we focus instruction on these observable tasks or behaviors, it is too easy to lose sight of this general goal.

George Madaus asked how Andy Porter's research related to the "conventional wisdom" in the testing field that test makers do a good job of sampling the material common to the major texts in a subject area. Andy related that the tests which were reviewed did, in fact, well assess the skills common to the three texts, but the material common to the texts represented only 7% of their content. On a related matter, David Cohen asked if the study's results had any implications for debates over whether a national curriculum now exists. At a high level of generalization, according to Andy, one could find evidence of such a curriculum across the texts which were reviewed: all elementary schools seem to teach children mathematics, and within this domain, all children are instructed in computational skills -- addition, subtraction, multiplication, and division -- and in measurement. However, below this level of abstraction, the curriculum, at least as it is represented in the texts which were analyzed, varies considerably. Holding the position that the transfer among skill domains is limited, Andy argued that teaching children to add fractions does not mean they will then know how to add decimals. For this reason, he does not see a national curriculum. Andy also noted -- this in response to a question from the Carnegie Corporation's Fritz Mosher -- that the generalizability of the findings were as yet limited. While he

and his colleagues have looked at math texts for grades 3 and 6 -- in addition to grades 4 and 5 -- the data from these analyses is not fully ready. However, some consistent differences among the three texts across the four grades seemed to be emerging (e.g., the Scott Foresman book consistently emphasized conceptual problems).

Marjorie Martus of the Ford Foundation asked Amity to comment on the policy implications, especially for teachers, of the assessment procedures she described. Amity felt the system is too new for its implications to be seen. This is the first year it is being used. Her major concern was that resources be found to implement the same process at other grades and to provide for staff development that needs to accompany the implementation of the system. She also noted that school principals are supporting the program. Les May of the Massachusetts Department of Education and later Dennis Gray of the Council for Basic Education all wanted to know about what safeguards existed to protect the Oakland procedure against teacher bias and even malice. Amity responded that the system on the one hand encourages trust among teachers by giving legitimacy to their judgments and at the same time provides specific criteria for them to use in making judgments. However, Amity acknowledged that the assessment procedure is not bias proof, but she believes that it may be less biased than more traditional assessments in that it provides teachers with a broad, rich data base upon which to make and justify instructional and promotional decisions.

Debby Meier offered that what teachers emphasize in terms of content and/or time may vary more with what is on tests than what is emphasized in texts. She noted along this line that parents and teachers often will view student difficulty on mathematics tests in terms of "basics" where in fact the source of the difficulty may be with high level skills. On a slightly

different matter, she discussed the difficulty of objectively testing "complexity." On sophisticated novels and movies, it is often hard to achieve agreement as to very basic matters such as the plot or the outcome. Debby believes that this sort of complexity leads test makers to focus on unchallenging, easy materials for comprehension tests. Following along this vein, Ann Cook related how one teacher justified the use of easy, uninteresting reading materials by noting that with Charlotte's Web even adults couldn't agree about the ending -- about whether or not it is a happy ending.

The issue of how difficult it is for adults to decide what something is about prompted Banesh Hoffmann to ask how teachers instruct children to select the right answers to main idea questions. Ann replied that if you are simply concerned that children perform well on the test, you teach them to translate the various guises of the main idea question into the question "What does this paragraph talk about most?". This strategy works most of the time. Later, however, Banesh gave an example of a paragraph for which the "talk most about" strategy nor any other formula strategy would work. In regard to the pervasiveness with which teachers employ test taking strategies, Andy Porter mentioned that considerable variation was found in a small survey his staff conducted of teachers' knowledge of tests.

Walt Haney raised the general issue of the "contextual dependency of test questions." He cited some examples from earlier in the conference to show that the meaning of a test item can shift depending on who is answering and the context in which it is given. Amity commented that the Oakland work was building in a context for assessment.

The impact of tests on teachers was raised by Frances Quinto of the NEA. She felt that if teachers are being evaluated on the basis of student test performance data -- which is apparently the case in some school systems -- then it is understandable and logical for them to stress test-taking strategies to their students. Andy noted that his study tried to assess impact of such accountability on teacher practice. Amity mentioned that in Oakland, teachers who actually stressed test strategies to children had less of an impact on their students' test scores than a group of teachers who were given the freedom to use the Teacher's Center to improve themselves in whatever way they wanted to. Judah Schwartz cited examples from the mathematical concepts sections of the MAT and the SAT to argue that test publishers and educators alike have a mechanical notion of mathematical concepts. He made a plea for a mathematics comprehension curriculum that stressed the understanding of the semantic referents for computational symbols in a manner akin to the understanding, in reading comprehension, of lexical symbols.

VI. REVIEW OF RECENT LEGISLATIVE PROPOSALS TO REGULATE TESTING

At the start of this session Andrew Strenio reiterated his thanks to Consortium members who had helped him in preparing his paper on open versus secure testing, which has been distributed as the sixth NCT staff circular. Walt Haney of Huron Institute, then provided a brief review of recent legislative proposals to regulate testing in various ways. He began by noting that there have been a host of developments along these lines since the last NCT conference in the spring of 1979. These fell into three categories: (1) state level proposals; (2) proposals introduced into the U. S. Congress; and (3) developments in New York in response to testing legislation signed into law in that state earlier this year.

At the state level, in addition to testing legislation enacted in California and in New York, similar proposals were introduced in Colorado, Maryland, Ohio, and Texas. However, proposals in none of these states apparently went very far. In addition, consideration has apparently been given to introduction of legislative proposals to regulate testing in Massachusetts, Michigan, Pennsylvania, Minnesota and Florida, although concrete information on what is or is not happening in some of these cases is hard to come by. Bills introduced in Maryland and Texas were apparently quite similar to the bill passed in New York. Rather than commenting in any detail on state level bills, however, Walt deferred to Rex Brown of ECS who is undertaking a detailed analysis of such state level proposals.

At the national level, two pieces of proposed legislation have been introduced into the U.S. Congress; one sponsored by Congressmen Ted Weiss of New York and George Miller of California, and by Congresswoman Shirley Chisholm of New York and another by Congressman Sam Gibbons of Florida. Copies of these proposed bills were previously distributed to all Consortium members and many have commented on them either in writing or in testimony before the House Subcommittee on Elementary, Secondary and Vocational Education which has been considering the proposed legislation. After six days of hearings on these proposals in the summer and fall, the subcommittee had planned markup of these bills in October. However, the scheduled markup has been postponed, in what one Congressional staffer called a "strategic withdrawal," because sponsors of the bills did not have sufficient votes to prevent their being killed.. Reportedly, sponsors plan to bring these proposals, back in spring 1980 for further consideration. Copies of testimony on these proposals, given in six days of hearings, are not yet available but are scheduled to be printed and ready for distribution within six to eight weeks.

Despite these developments, Walt commented, the focus of attention of legislative initiatives to regulate testing remains in New York. The New York bill which was the focus of one panel at the spring 1979 NCT conference was enacted into law this past summer and the New York State Department of Education has been given responsibility for implementing the law. So far the Department is reported to have undertaken two activities with regard to this implementation. First they surveyed test publishers to find out which tests publishers judged to fall within the purview of the legislation. On that basis twenty-six (26) testing programs were identified as being directly affected by the new legislation. In addition, the Department asked publishers to indicate how they would respond to the law's requirement of test disclosure after administration. As the Department reported in an initial impact report on the law (see summary, attachment 9), "Based on very tentative information available by October 15 [1979], about twenty of these programs may be withdrawn from use in the state." Walt went on to summarize the findings reported by the New York State Department of Education, as reported in its initial impact report of Oct. 16, 1979 [see attachment 9]. According to this account six major admissions testing programs likely will continue in New York under the provisions of the new law -- five published by the Educational Testing Service for various sponsors and the sixth being the ACT Assessment Program sponsored by the American College Testing Program.

As a result of these findings Walt recounted, consideration is being given in New York to the possibility of exempting some smaller volume tests from the disclosure provision of the new law, but as far as is currently known that possibility has not been pursued in concrete form.

At this point, Walt asked Rex Brown of the Education Commission of the States to describe work he is planning to analyze legislative proposals on

testing. Rex recounted that he and Merle McClung were preparing a report which will likely draw on some of Andy Strenio's analysis of current legislation, but which will be aimed first of all at ECS's primary constituency, namely state legislators who are likely to have to deal with similar proposed legislation on testing. So, the ECS report will attempt to lay out some of the issues which would come up in legislative testimony on proposals to regulate testing, regarding standardized testing in general, admissions testing in particular, regarding government regulation of this industry and security versus openness in tests. As part of the ECS report, Rex continued, they will try to describe all of the legislative proposals which can be found. Rex also mentioned that he has heard that bills to regulate testing likely will come up again in Florida and in Maryland. The ECS effort will attempt to analyze commonalities and differences among state proposals and the Weiss and Gibbons bills at the federal level, and to review them in terms of legal issues, Constitutional issues, issues of restraint of commerce, copyright and things like that. He hopes that by the time the report is finished more will be known about the impact of the New York law. Nevertheless, it is hoped that the ECS study will be completed within the next couple of months. Rex mentioned too that the proposals to regulate testing, often reflect other broader concerns, so that people on opposite sides of these proposals are often not, as Andy pointed out, connecting in terms of their arguments. If people have direct experience with these proposals or views on what they are really about, Rex invited them to pass along their ideas to him at ECS (Rex Brown, Education Commission of the States, Suite 800, Lincoln Tower, 1860 Lincoln St., Denver, Colorado 80295, Phone (303) 861-4917).

One person from the audience asked Rex what the ECS position on legislative proposals to require open testing would be in light of the fact that ECS sponsors the National Assessment of Educational Progress which already does release at least some questions after administration. Rex said that the more he looks into the legislative proposals to regulate testing the more he realizes the complexity behind them, and hence the difficulty of taking simple positions on them. He said, however, that the report he is preparing may provide a basis for ECS through its governing council to take some stand on these issues.

Laura Hines of the APA reiterated Rex's concern about the complexity of the issues involved in the proposals to regulate testing. She noted that the APA had asked Governor Carey to veto the New York bill, not because it was necessarily opposed to much of the intent behind the bill, but simply because there had not been sufficient time to consider all the complexities involved.

In response to a question from Frances Quinto, Rex explained that in December ECS would be holding a mock legislative hearing on an open testing bill simply as a means of familiarizing legislators with the issues involved in such proposals. Vito Perrone and someone from ETS would participate to present different perspectives on the mock bill to be considered. Vito noted that he would participate in the mock hearing not as a representative of the NCT, but simply to represent his own views.

Another questioner asked Walt when the New York law was to take effect. He said that legally it goes into effect on January 1, 1980, but that since several early 1980 test administrations had been cancelled, he was uncertain

as to when the first test would actually be released after administration. At this point Walt asked Alice Irby of ETS if she could provide more information on this point since ETS was the publisher of five of the six admissions tests which apparently would continue to be given in New York. Alice explained that the five tests to which Walt referred are the GMAT, GRE, LSAT, SAT and the TOEL. She noted, however, that there would be some reduction in services offered with respect to these tests. Also she briefly summarized several proposals made by Robert Solomon of ETS in testimony before Congress. These included proposals for the National Academy of Sciences Committee on Ability Testing to look into the issues under debate, for the American Council on Education to appoint a committee on test use to provide guidance to colleges and universities regarding test use, for ETS to sponsor a special conference on test legislation and also for the public release of one of the forms of the National Teacher Exam in either 1980-81 or 1981-82. With regard to the latter proposal, ETS has invited the AFT, the NEA and the AACTE (American Association of Colleges of Teacher Education) to cooperate with ETS to prepare materials to accompany release of the NTE form.

VII. THE TEST COACHING CONTROVERSY: WHAT THE FTC REPORTS DID AND DID NOT SAY

In a session following up on a panel at the spring 1979 conference, Walt Haney provided a review of the recently released Federal Trade Commission reports concerning test coaching and preparation schools. He began by noting that in describing the FTC reports he was substituting for two people who have studied the test coaching and preparation issue in considerable detail, but who could not attend the NCT conference. Henry Dyer who published one

of the first studies of test preparation in the 1950's could not attend the conference due to illness. Lewis Pike, author of one of the most complete reviews of the test coaching and preparation literature was not able to attend due to unforeseen circumstances.

Walt began his review of the FTC reports by noting that having two versions of the FTC study of test coaching schools helps nicely to prevent what Decker Walker worried about, namely oversimplification of public policy issues. He mentioned also that summaries of both reports had previously been distributed to NCT members.

The first report, by the Boston regional office of the FTC, dated September 1978, provides: (1) background to the FTC study; (2) a history of the SAT, the LSAT and their uses; (3) review of related research and literature; (4) description of investigative methodology employed by the FTC; (5) a review of the structure of the test coaching and preparation industry; (6) a description of the statistical and computer methodology used in the Boston study; and (7) findings from the study.

In contrast, the Washington FTC report, dated March 1979, focuses almost exclusively on data analysis issues. It does not attempt to repeat the earlier report's review of the literature, description of industry structure, or history of the SAT, LSAT and their uses. Hence, Walt suggested, despite the fact that much public attention has focused on the conflicting findings of the two FTC reports, they do not actually cover the exact same territory, and hence each contains some very interesting information not covered in the other.

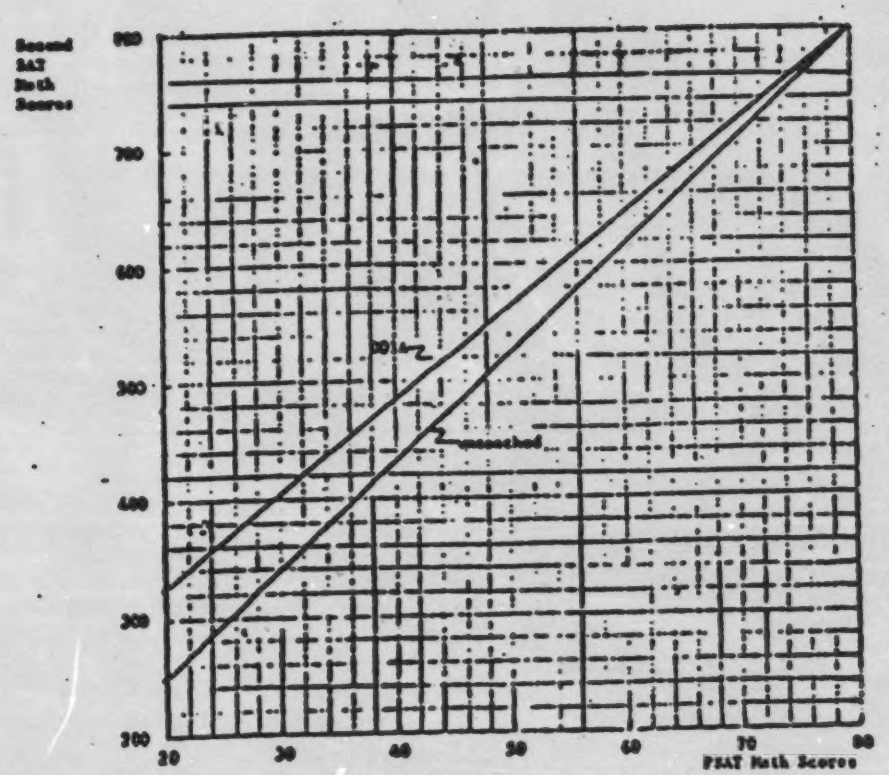
Next, Walt explained the differences in methodology employed in the Boston and Washington studies to derive estimates of the effects of test coaching. To illustrate the methodology used in the Boston study he showed

an overhead transparency of a scatterplot of scores which students received on the PSAT and the SAT. (see Figure 1) From such distributions, regression lines were calculated separately for students who did not attend a coaching school and for students who did attend a coaching school between taking the PSAT and the SAT (or between first and second administration of the SAT or the LSAT). The Boston study interpreted the difference in height between these two regression lines as the effect of test coaching. The fact that the two lines are farthest apart at the lower end of the scale illustrates the basis on which it was concluded that coaching is most effective for "underachievers". In response to a comment from Carlyle Maw, Walt acknowledged that in one respect this approach may underestimate test coaching effects, since some of the students included in the "uncoached" group may have actually received coaching even if it was not in commercial coaching schools from which the coached students were identified. In response to a question from Eric Gardner, Walt commented that the issue of self-selection of students into coaching schools was not directly addressed in the Boston study, but that it was treated in the Washington report. In response to another question, it was explained that the magnitude of effects of coaching estimated in the Boston study ranged from around 75 points on the lower end of the SAT scale to zero at the upper end of the scale.

The Washington office of the FTC made three major criticisms of the Boston study:

1. The Boston study failed to control for differences that might exist between coached and uncoached students in terms of personal and demographic variables.

Figure 1: Scatterplot Showing Separate Regression Lines for Coached and Uncoached Students (From Boston FTC Report p. 82).



SAT average math scores for second-SAT as a function of PSAT math scores, comparing 3-time takers coached at CBSA before their first SAT against uncoached 3-time takers

2. The Boston study failed to include tests of statistical significance for the effects estimates reported.

3. The Boston study improperly attempted to extrapolate findings beyond the range of available data, particularly with respect to findings regarding coaching on the LSAT.

Because of these criticisms, the Boston report was issued by the FTC Washington office with the disclaimer that it "has several major flaws in the data analysis, making the results unreliable."

In its own study the Washington office of the FTC, specifically the Bureau of Consumer Protection, reanalyzed available data using a more sophisticated analytical technique. Specifically it used a multiple regression analysis in which it attempted to adjust for a variety of background variables on which coached and uncoached students differed. These included:

- Rank in high school class;
- Sex;
- Grade in last English (for verbal SAT) or Math (for math SAT) course;
- Parental income;
- High school type -- public or nonpublic (many of the non-public schools in the market area are parochial schools);
- Years of English/Math expected to be taken by high school graduation;
- Score on PSAT (a preliminary exam taken before the SAT);
- Number of PSAT's taken;
- Elapsed time from first PSAT to SAT;

On the basis of this reanalysis, the Washington office derived estimates of effects of test coaching which were much more modest, but which in most cases were still significant at least for the school identified as school A, namely the Stanley Kaplan Educational Center. For this school, estimates of effects for various subsamples (students taking the SAT on different dates for a first or second time, and pooled samples) were:

Adapted From TABLE 3 of Washington FTC Report

OVERALL IMPACT OF COACHING

Mean Number of Points (and Confidence Interval)
Contributed by Coaching School

		School A	
	Subsample	Verbal SAT	Math SAT
1.	1st SAT 4/75 (n=476)	18.0 (2 - 34)	17.1 (0 - 34)
2.	1st SAT 4/76 (n=658)	44.5 (33 - 57)	26.5 (14 - 39)
3.	2nd SAT 11/76 (n-359)	26.9 (11 - 43)	22.4 (5 - 40)
4.	2nd SAT 11/76 (n=438)	25.4 (14 - 36)	30.7 (19 - 42)
5.	1st SAT - Pooled Time Periods (n=1578)	29.7 (21 - 39)	19.2 (10 - 28)
6.	2nd SAT - Pooled Time Periods (n=1176)	27.2 (19 - 35)	28.4 (20 - 37)

In response to a question from Alice Irby, Walt pointed out that it was only for this one school of the two studied, for which significant effects estimates were derived. With respect to this finding, Walt commented that he had not seen Stanley Kaplan quoting these findings in advertisements. In response to this observation Stanley Kaplan said that he did not believe the findings show very much. He pointed out the second school mentioned in the Washington FTC report was actually misidentified. He suggested that somewhere along the line the report loses credibility when it identifies

the wrong school.

In response to a question from George Madaus, Walt explained that the sample sizes on which the FTC analyses were based ranged widely across different subsamples, from only 300-400 to around 1500 for subsamples for the SAT. However, for the LSAT sample sizes were smaller and it was apparently for this reason that the Washington report devoted no attention to analyzing the LSAT scores.

Walt continued by explaining that the Washington study also attempted in one reanalysis to control further for the possible effects of self-selection of students in the coaching schools and found that results from this reanalysis indicated smaller but in half the cases still significant effects estimates for school A.

While most of the attention concerning the FTC studies has focused on such effects estimates, Walt pointed out, both of them also contain some interesting descriptive information. Background information on coached and uncoached students, for example, suggests that these groups do not differ markedly on sex and ethnicity, but do differ substantially on type high school and parental income, as shown in following data;

Adapted from TABLE 1 of Washington FTC Report

COMPARISON OF DEMOGRAPHIC PROFILES OF COACHED AND UNCOACHED STUDENTS
(page 1)

HIGH SCHOOL TYPE**	Coached	Uncoached
Public	55.3%	75.4%
Other then Public	44.7%	24.6%
	100.0%	100.0%
	(n=988)	(n=1706)
PARENTAL INCOME		
Less than $12,000 per year	15.7%	23.3%
Between $12,000 and $17,999	15.6%	25.9%
Between $12,000 and $23,999	16.2%	20.5%
Between $24,000 and $29,999	11.3%	13.0%
$30,000 or more	41.2%	17.2%
	100.0%	100.0%
	(n=770)	(n=1395)

** Coached and uncoached groups are significantly different at the .01 level.

In response to a question, Walt noted that these data were drawn from student reports and were not, in any case, necessarily generalizable to broader populations of students who attend various types of coaching schools.

He also noted that the FTC study uncovered the fact, recounted in the Boston office's report that one commercial coaching enterprise actually pays students to take secure tests for the sole purpose of memorizing answers on the test. Also one university sponsors a coaching course for fees comparable to those of commercial coaching schools, and at the same time relies upon results of tests for which it provides coaching in its admissions process. Regardless of the empirical question of whether or not the university's coaching is effective, Walt suggested, such practices would seem to raise important educational policy issues.

In summary, Walt suggested, there is considerable descriptive information in both FTC reports besides the estimates of effects of the two coaching schools studied. Hence, it would be worth the while of anyone interested in these issues to have a look at both of them. Walt's presentation was followed by comments and questions from the audience.

Frances Quinto of the National Education Association (NEA) mentioned that the NEA is continuing its interest in the test coaching issue, and is now taking depositions from FTC officials concerning the possibility of gaining access to certain information which the FTC obtained during its study of coaching schools, but which is now considered to be privileged information.

If test coaching is effective, asked Deborah Meier, what does this mean? What should we do about it? Walt said that this is a question which he had hoped Henry Dyer and Lew Pike would address, but he nevertheless went on to mention a few implications which had been drawn from the FTC reports or which ought to be raised. The Washington FTC report suggested that further inquiry

on the effects of coaching is needed. Others have suggested that adjustments to students' scores might be made in light of whether or not students had received coaching. This proposal would, however, in Walt's view, seem to lead to some rather Byzantine procedures. Another question clearly raised by the FTC findings is the perceived value we place on such test scores as predictors of how well students will do in college. Unfortunately, however, neither of the FTC studies looked into the possible effects of coaching on the predictive validity of these admissions tests.

Walt suggested also that it might be partially due to such findings that colleges apparently are placing decreased emphasis on admissions best scores in the college admissions process -- although, he went on, this change is probably more likely to be due to demographic changes in the college-going population.

Banesh Hoffmann commented upon a recent incident concerning another standardized admissions test. Referring to an article in the Oct. 27, 1979 New York Times, he recounted that the incident concerned a nursing exam taken by thousands of individuals. Anonymous letters were received that cheating had taken place on the exam. Education officials initially decided to throw out the results, but they reversed themselves and instead requested a special investigation of the issue. The investigation concluded that there was no evidence that any part of the examination was available in advance to any candidate or other unauthorized person and that there was no evidence of misconduct by anyone involved. Instead the investigation found that the drastic improvements in scores -- as much as 400 points out of a possible 800 from one administration to the next, which was part of the basis on which

officials originally had decided to discard the results -- were caused by candidates' participation in examination review courses. Interviews with some of those who had allegedly cheated found that all of the candidates interviewed had participated in at least one examination review course." So, Banesh commented, officials have decided not to cancel the scores on the grounds that these review courses produced the remarkable gains in students' scores.

In response to a question from George Madaus, Walt explained that neither of the FTC studies had addressed the issue of test coaching effects in terms of unique and overlapping variance explained.

Vito Perrone asked Stanley Kaplan about the degree to which his business might have increased as a result of the FTC inquiries. Stanley Kaplan commented that his enrollments increased very largely as a result of the publication of the Wirtz SAT report. He also remarked that the FTC had never challenged the coachability of admissions tests like the GMAT, but had only addressed the SAT and the LSAT which were viewed as aptitude tests. Continuing, Stanley observed that the FTC found no evidence of effectiveness for the coaching school which has students take secure tests for the purpose of memorizing questions. This illustrates, he said, what he is afraid of when it comes to truth-in-testing, namely that people will see released questions, memorize them, and the whole purpose of the testing program will be frustrated.

Bernard McKenna suggested that if coaching schools are effective, perhaps we ought to pay for all students to attend these schools, so that not just the rich can afford them.

Another person commented that the terms employed in discussing coaching issues tend to color the debate on them. If, for example, we use the terms

cheating or special instruction, we would likely view the issues very differently. In this respect he noted that we already do pay to prepare all students to pass these tests -- it's called public education. But there are people who are not satisfied with that form of coaching, and who don't go to Mr. Kaplan's course. Rather they go to Andover, it was suggested, which can indeed be viewed as a coaching school.

To this suggestion, Walt offered two observations. First is that although private schools might be viewed as coaching schools, it should be noted that according to the FTC data, non-public school students attend commercial coaching schools at a much higher rate than do public school students. Second, the crucial issue in this debate is what learning takes place as a result of coaching versus what learning results from schooling.

To resolve this issue, Walt suggested, what we would need would be some independent criterion of learning besides the admissions tests to see whether coaching simply boosts admissions test scores or whether it also fosters broader learning. Unfortunately, however, Walt said that he knew of no research in the test coaching literature which addresses effects of coaching on outcomes besides the admissions tests.

Deborah Meier suggested that teachers try to deal with this dilemma by asking whether they are teaching in the same way as they would if these tests did not exist. In this respect, she maintained the tests clearly are affecting the educational process. Things on tests get stressed which otherwise would not be stressed. Continuing, with an example about instructing students as to whether or not to guess on tests, she said that it is often very hard to determine exactly where legitimate test preparation leaves off and cheating begins.

VIII. SUMMARY OF SMALL GROUP SESSIONS

Following lunch on Tuesday, brief summaries were given of the small group sessions which had taken place that morning.

1. Plans for Case Study of South Carolina Basic Skills Assessment Program

This session was summarized by Walt Haney. He recounted that discussion of this session centered first on the nature of the South Carolina Basic Skills Assessment Program. This discussion was facilitated by the participation of Vana Meredith of the South Carolina State Department of Education. The South Carolina endeavor is not a competency testing program. Rather it is a basic skills assessment program. It aims at helping to improve instruction for students who are weak in the basic skills. These are defined in terms of basic skills objectives which have been identified through an extensive process involving state officials, local district personnel, teachers and the public generally. Eighteen objectives have been defined at the kindergarten level and sixteen objectives have been set for grades 1-12. The assessment program started this fall with implementation of a program to screen youngsters upon entry into first grade with an instrument called the Cognitive Skills Assessment Battery. It was selected by officials in South Carolina because it represented a relatively good match with the kindergarten objectives previously established. The intention of the program is explicitly non-punitive, that is, it is not intended to control grade-to-grade promotion. The assessment program will be implemented gradually over the next several years into the higher grade levels. In light of this schedule, the NCT case study of this program will be able to examine actual implementation of the South Carolina Basic Skills Assessment program only at the school entry level.

After description and discussion of the program itself, the small group

session focused on the key issues which the case study ought to address. The most general such issue identified was the educational impact of the assessment program -- namely how education in South Carolina actually is changing as a result of the program. The second issue raised in the small group session was how the assessment program is actually being presented and described to the public, and what the local perceptions of the program actually are, whatever the intentions and perceptions of state officials may be. Specifically people inquired as to what local perceptions, by citizens, teachers and children actually are.

Another question raised in the small group session was exactly what has been the nature of citizen and parent involvement in the development of the program, and also how they will be involved in the instructional response to the assessment results, a response which will have to be forthcoming at the local level.

Another concern expressed in the session was whether the assessment program, with standard assessments made in first, second, and third grade, will be sufficiently flexible to provide for a developmental rather than a static perspective on children's development.

A final concern expressed by several people was that the panel constituted to carry out the case study encompass several of the diverse perspectives represented in the Consortium; for example those of teachers, minorities, parents, educational administrators, and experts in educational testing and measurement.

2. Plans for Case Study of New York Citywide Reading Test

An NCT inspired case study of New York City's Citywide Reading Achievement Testing Program was discussed in another of the small sessions held on Tuesday morning. Gene Radwin of the Huron Institute, who is coordinating this project, outlined the goals and nature of this Ford Foundation supported work. In brief,

this case study is designed to examine and offer suggestions for improving NYC's annual testing program in reading, for all children in grades 2 - 8. To develop recommendations which will be useful to the New York City school system (and to other large urban school districts which conduct system-wide testing programs), the Huron staff and a small panel nominated by NCT members -- which will represent different perspectives on testing -- will carefully consider the functions of the testing and of how well these functions are met.

At several recent NCT conferences, Ann Cook and Debby Meier had brought attention to the New York City program, by discussing such issues as the uses which the testing serves and controversies over the security and difficulty of the test instrument. A series of events in NYC last spring and summer (some of which were discussed at the spring 1979 NCT conference) served to motivate Huron staff for the Consortium to undertake, and the Ford Foundation to underwrite, the current case study.

According to Gene and to NCT conference participants familiar with this story, the spring 1979 administration of the Citywide Testing Program, which employed the newest version of the Psychological Corporation's Metropolitan Achievement Test [MAT], became ensnarled in considerable controversy. After discrepancies were discovered in the test results, for which the city was unwilling to accept remedies proposed by the MAT publisher, it was decided to retest, in early June, all the children in New York City with the California Achievement Test (CAT). One of the local districts in New York refused to participate in this testing, arguing that reason had not been demonstrated for a retest. Later, when the state department of education accepted the MAT scores from this district for the sake of Title 1 evaluation, it gave tacit approval to this district's action.

The choice of the CAT for the retest was to prove significant. The CAT and the MAT were the only two tests to have been approved by the Board of Education's test selection committee for the citywide norm referenced testing program. The CAT was also one of three tests approved for the newly designed diagnostic testing program that was to begin in the fall of 1979. To the Board of Education's embarrassment, however, the CAT was the only test to compete for the diagnostic program contract. Thus, New York City was faced with the situation in which a test that had just been given as a norm referenced achievement instrument was to be administered several months later as a diagnostic/criterion referenced instrument. (McGraw-Hill, the publisher, does provide different analyses of the results, depending on which purpose the test is meant to serve.) To forstall further controversy, the Board of Education used the results from the June administration of the CAT to obtain diagnostic analyses of pupils' needs. However, as the scores for many children were lost in their transition from one grade to the other, and as new children were entering the system, many children did have to be tested in the fall.

In this jumble of events, Gene recounted, the key issues that surround the citywide testing program can be discerned. Among the most important of these issues are the purposes which the testing is intended to serve. Two main reasons for the testing, and specifically for the retesting, are that the testing is mandated by state law and that it is needed for the evaluation of federally funded programs. The law decentralizing the city's school system requires that schools be ranked in terms of pupil performance on the citywide test and then permits those schools whose pupils score most poorly to follow special certification standards for selecting teachers. Ann Cook noted, however, that for the

last several years the city has not had the money to hire any teachers and that a norm referenced instrument may not be necessary for the evaluation of federally funded programs.

Along this vein, another stated purpose of the testing is to help teachers make promotional and placement decisions. However, current practices raise questions about how this purpose is served. According to central school administration guidelines, the results on the citywide should lead to 40% of the children in New York City being held back every year. In fact, only 8% of them are. In spite of widespread dissatisfaction with the testing program, Gene suggested that to some extent both professional groups and parent groups view the test as providing themselves with a sort of protection. The parent groups do not trust teacher judgment as a sufficient basis for making promotional and placement decisions. And the professional groups want to have test scores to support their decisions.

Another issue which emerges from the events last spring is the logistics of the testing program. Testing five hundred thousand children yearly is an enormous task. The materials needed for the testing -- test booklets, answer booklets, teachers manuals, etc. -- could, according to one board of education official, cover a football field. To administer the test is therefore an enormous task requiring considerable preparation. To retest a half million students in only three weeks time as New York City did last spring, whatever else one may say, was an administrative achievement of no small scale. Thus, any consideration of the citywide testing program must consider logistical questions along with more substantive and procedural ones.

In addition to understanding the purposes of the testing, Gene said that

the test selection process needs to be considered. Testing five hundred thousand students involves a considerable amount of money. Several years ago, questions were raised about possible conflicts of interest between New York City board of education officials and the publisher of the test used in the citywide program. In response to such questions, the board now employs a selection committee composed of administrators, teachers, parents and central board officials to screen potential tests. After an initial screening, tests are considered on a competitive basis; the publisher who is the lowest bidder is selected. As a result of this process, several different tests have been used over the last few years. Dissatisfaction with continuing changes in the test used has led to consideration of other means of test selection. More germane to what occurred last spring, is the issue raised by NCT members of why the test selection committee or central board staff did not perceive the problem with the MAT prior to its administration.

Another issue which emerges from events last spring is the entangling relationships of federal, state, city, and local education agencies in regard to the city's testing program. As noted previously, the test is mandated by state law. Moreover, the state established requirements for which tests can be used to evaluate federally funded programs. Also the state requires the central board to administer several other testing programs. In addition, the state and the local school districts can often bypass the city in regard to testing issues. Thus, it was possible to have a situation last spring in which the district that defied the city (by not participating in the retesting) could have its actions implicitly sanctioned by the state.

Test security, preparation, and cheating are also important issues within the New York testing program. Teachers and principals are under pressure to have their pupils do well on the test. Schools' test results are printed in

the New York newspapers. Also the current director of the Educational Evaluation Office has maintained that these test results can be used for teacher evaluation. According to Ann and Debby and officials Gene interviewed, cheating on the test is pervasive. On several occasions reports of widescale cheating on the test have hit the media. Two years ago teacher and parent groups brought an injunction against administering the test, arguing that its security had been undermined by the distribution of copies of the test in some districts prior to administration.

In addition to these and other issues which emerged from Gene's interviews with people in New York City, people participating in the small group session raised other issues for consideration in the case study. For example, Judah Schwartz raised the question of whether aggregate data, such as is obtained from the type of testing program the city employs, can be used to make decisions on individual children. George Madaus expressed concern that the case study consider the size of central board's testing program, the number of individuals involved, the budget, etc. He was also interested in having consideration given to the city developing its own tests, instead of relying on commercially developed ones. Others urged that Gene and the panel for the study not accept at face value official views about the testing program. The need to dig below the surface of the test program was stressed by several people.

3. Discussion of Possible NCT Collaborative Study of Test Use and Development of Materials on Testing

A third small group session on Tuesday morning dealt with possible collaborative undertakings by NCT members to conduct studies of test use or to develop and share materials on testing for public education. Vito Perrone and David Cohen chaired this small-group session. Vito opened the session by noting that

it harkened back to one of the early goals of the Consortium, namely to improve communication and cooperation among Consortium members. He suggested that member organizations might share publications information so that information developed by one organization might be shared more widely. Also discussed was the possibility that even if particular topics were not of interest to the NCT membership as a whole, smaller subsets of Consortium members might collaborate in treating them.

The prime concern in the small group discussion was the utilization of tests and evaluation information, and ways in which utilization concerns might be focused so that NCT resources were better used. Ann Kahn of the PTA expressed interest in developing materials which would help parents delve into testing issues in their own communities. While such materials would help to inform a parent perspective on testing, if they were developed in collaboration with other NCT groups, they might also inform other perspectives, for example those of teachers, administrators and students. Eva Baker expressed interest in a study of the sort Ann Kahn mentioned, and briefly described a study of test utilization by the Center for the Study of Evaluation. She suggested that some members concerns might be reflected in that study. Bernard McKenna of the NEA described a teacher training activity which the NEA has undertaken to improve teachers' awareness and skill in respect to testing issues, and offered to make the materials available to others interested.

Vito Perrone noted that it is important not only to determine what people's perceptions on testing are, but also how their perceptions get changed. He expressed his personal enthusiasm for getting involved in some sort of collaborative venture regarding utilization. It was agreed that the small-group discussion on this point concluded with the expressed

intention that NCT staff at Huron should serve as a sort of clearinghouse in promoting such collaborative ventures, and in connecting organizations with similar interests. One step in this direction was a quick meeting of those interested in the testing utilization issue at the end of the discussion. Another step will be the circulation of these notes.

Concerns were also expressed for improving communication among Consortium members. Vito noted that occasionally the NCT has sent out pieces of paper soliciting information or reactions of various sorts from member organizations, but generally responses to such requests have not been good. In this respect, Vito commented that perhaps we ought to attempt again such inquiries since now the NCT has a longer history of work and at least some collaborations. Perhaps this would serve as one means of bringing organizations with common interests together in collaborative endeavors.

Walt Haney will put together a sort of directory to NCT member organizations, and asked member organization representatives to fill out the forms which he had enclosed in representatives' conference packets. Also, both Vito and Walt invited representatives to nominate individuals to serve on the NCT case study panels.

On that note Vito brought the general session of the fall 1979 membership conference to a close. The NCT business meeting followed. An account of that meeting has been prepared and distributed separately to NCT members.

SPEECH COMMUNICATION ASSOCIATION
FOUNDED 1914

National Office
5205 LEESBURG PIKE
FALLS CHURCH, VA 22041
(703) 379 1888

CRITERIA FOR EVALUATING

INSTRUMENTS AND PROCEDURES FOR

ASSESSING SPEAKING AND LISTENING*

The following criteria may be applied to published and unpublished instruments and procedures for assessing speaking and listening skills of children and adults. The criteria are organized around (a) content considerations, which deal primarily with the substance of speaking and listening instruments and procedures, and (b) technical considerations, which deal with such matters as reliability, validity and information on administration.

1. Stimulus materials should require the individual being tested to demonstrate skill as a speaker or listener.

2. Assessment instruments and procedures should clearly distinguish speaking and listening performance from reading and writing ability; i.e., inferences of speaking and listening competence should not be made from tests of reading and writing, and directions and responses for speaking and/or listening tests should not be mediated through reading and writing modes.

3. Assessment instruments and procedures should be free of sexual, cultural, racial, and ethnic content and/or stereotyping.

4. Assessment should confirm the presence or absence of skills, not diagnose reasons why individuals demonstrate or fail to demonstrate those skills.

5. Assessment should emphasize the application of speaking and listening skills that relate to familiar situations, i.e., stimulus materials should refer to situations recognizable to the individual being tested and should facilitate demonstration of skills rather than demonstration of content mastery.

6. Assessment should test skills that are important for various communication settings (e.g., interpersonal, small group, public, and mass communication settings) rather than be limited to one setting.

7. Assessment should permit a range of acceptable responses, where such a range is appropriate.

*Developed by Philip M. Backlund, Kenneth L. Brown, Joanne Gurry, and Fred E. Jandt acting as a subgroup of the Speech Communication Association's Educational Policies Board Task Force on Assessment and Testing. Approved and endorsed by the Educational Policies Board and the Administrative Committee of the Speech Communication Association.

8. Assessment should demonstrate that outcomes are more than just chance evidence; i.e., assessment should be reliable.

9. Assessment should provide results that are consistent with other evidence that might be available.

10. Assessment should have content validity.

11. Assessment procedures should be standardized and detailed enough so that individual responses will not be affected by the administrator's skills in administering the procedures.

12. Assessment procedures should approximate the recognized stress level of oral communication; they should not increase or eliminate it.

13. Assessment procedures should be practical in terms of cost and time.

14. Assessment should involve simple equipment.

15. Assessment should be suitable for the developmental level of the individual being tested.

ATTACHMENT II
Statement sponsored by twelve organizations: "The Essentials of Education"

THE ESSENTIALS OF EDUCATION

Preamble

Society must reaffirm the value of a balanced education.

Leaders of several professional organizations reached this conclusion in 1978. They circulated a statement on the essentials of education among a number of professional associations whose governing boards endorsed the statement and urged that it be called to the immediate attention of the entire education community, of policy makers and of the public at large.

The statement that follows embodies the collective concern of the endorsing associations. It expresses their call for a renewed commitment to a more complete and more fulfilling education for all.

The associations invite the concurrence, support, and participation of everyone interested in education.

Public concern about basic knowledge and the basic skills in education is valid. Society should continually seek out, define, and then provide for every person those elements of education that are essential to a productive and meaningful life.

The basic elements of knowledge and skill are only a part of the essentials of education. In an era dominated by cries for going "back to the basics," for "minimal competencies," and for "survival skills," society should reject simplistic solutions and declare a commitment to the essentials of education.

A definition of the essentials of education should avoid three easy tendencies: to limit the essentials to "the three R's" in a society that is highly technological and complex; to define the essentials by what is tested at a time when tests are severely limited in what they can measure; and to reduce the essentials to a few "skills" when it is obvious that people use a combination of skills, knowledge, and feelings to come to terms with their world. By rejecting these simplistic tendencies, educators will avoid concentration on training in a few skills at the expense of preparing students for the changing world in which they must live.

Educators should resist pressures to concentrate solely upon easy-to-teach, easy-to-test bits of knowledge, and must go beyond short-term objectives of training for jobs or producing citizens who can perform routine tasks but cannot apply their knowledge or skills, cannot reason about their society, and cannot make informed judgments.

What, then, are the essentials of education?

Educators agree that the overarching goal of education is to develop informed, thinking citizens capable of participating in both domestic and world affairs. The development of such citizens depends not only upon education for citizenship but also upon other essentials of education shared by all subjects.

The interdependence of skills and content is the central concept of the essentials of education. Skills and abilities do not grow in isolation from content. In all subjects, students develop skills in using language and other symbol systems; they develop the ability to reason; they undergo experiences that lead to emotional and social maturity. Students master these skills and abilities

through observing, listening, reading, talking, and writing about science, mathematics, history and the social sciences, the arts and other aspects of our intellectual, social, and cultural heritage. As they learn about their world and its heritage they necessarily deepen their skills in language and reasoning and acquire the basis for emotional, aesthetic and social growth. They also become aware of the world around them and develop an understanding and appreciation of the interdependence of the many facets of that world.

More specifically, the essentials of education include the ability to use language, to think, and to communicate effectively; to use mathematical knowledge and methods to solve problems; to reason logically; to use abstractions and symbols with power and ease; to apply and to understand scientific knowledge and methods; to make use of technology and to understand its limitations; to express oneself through the arts and to understand the artistic expressions of others; to understand other languages and cultures; to understand spatial relationships; to apply knowledge about health, nutrition, and physical activity; to acquire the capacity to meet unexpected challenges; to make informed value judgments; to recognize and to use one's full learning potential; and to prepare to go on learning for a lifetime.

Such a definition calls for a realization that all disciplines must join together and acknowledge their interdependence. Determining the essentials of education is a continuing process, far more demanding and significant than listing isolated skills assumed to be basic. Putting the essentials of education into practice requires instructional programs based on this new sense of interdependence.

Educators must also join with many segments of society to specify the essentials of education more fully. Among these segments are legislators, school boards, parents, students, workers' organizations, businesses, publishers, and other groups and individuals with an interest in education. All must now participate in a coordinated effort on behalf of society to confront this task. Everyone has a stake in the essentials of education.

American Council on the Teaching of Foreign Languages
2 Park Avenue, Room 1814, New York, New York 10016
(212) 689-8021

Association for Supervision and Curriculum Development
225 N. Washington Street, Alexandria, VA 22314
(703) 549-9110

International Reading Association
800 Barksdale Road, Box 8139, Newark, Delaware 19711
(302) 731-1600

Music Educators National Conference
1902 Association Drive, Reston, Virginia 22091
(703) 860-4000

National Art Education Association
1916 Association Drive, Reston, Virginia 22091
(703) 860-8000

National Association of Elementary School Principals
1801 N. Moore Street, Arlington, Virginia 22209
(703) 528-6000

National Council for the Social Studies
Suite 400, 2030 M Street, N.W., Washington, D.C.
20036. (202) 296-0760

National Council of Teachers of English
1111 Kenyon Road, Urbana, Illinois 61801
(217) 328-3870

National Council of Teachers of Mathematics
1906 Association Drive, Reston, Virginia 22091
(703) 620-9840

National Science Teachers Association
1742 Connecticut Avenue, N.W., Washington, D.C. 20009
(202) 265-4150

Speech Communication Association
5205 Leesburg Pike, Falls Church, Virginia 22041
(703) 379-1888

ATTACHMENT III
Description of the Math/Science Network

WHY ARE WOMEN:

1% of auto mechanics
2% of secondary school principals
2% of engineers
2% of apprentices in the skilled trades
3% of college or university presidents
4% of police officers
9% of doctors
15% of life and physical scientists
99% of the secretarial work force . . . ?

CONSIDER THESE PATTERNS OF EDUCATION AND SOCIALIZATION . . .

In Elementary School

Girls and boys like math and science equally, and their achievement scores are identical, yet . . .

During High School

Females take fewer math, science, and industrial arts classes than males, and . . .

Over 50% of female students enrolled in vocational programs are being trained for homemaking, and one-third for clerical work.

After High School

Women are concentrated in clerical and service jobs, or work as teachers, nurses, and librarians.

The average woman worker earns less than three-fifths of what a man does, even when both work full-time, year-round.

IT'S TIME TO CHANGE

Current educational training limits the options of young women and confines them to a narrow range of occupations.

Expanding options for girls and women is the purpose of

THE MATH/SCIENCE NETWORK

This association of individuals and organizations was established to promote the participation of women in mathematics and science, and to encourage their entry into nontraditional occupations.

Network members are scientists, educators, engineers, parents, community leaders, business and industry people.

NETWORK MEMBERS:

* create conferences
* provide technical assistance
* train teachers
* produce film and print materials
* act as a forum for discussion
* develop school activities
* provide speakers
* promote community education
* share employment information

THE MATH/SCIENCE NETWORK is supported by a grant from the Carnegie Corporation of New York and coordinated through Mills College (Oakland, California) and the Lawrence Hall of Science (University of California, Berkeley).

IF YOU ARE . . .

an educator concerned with increasing female participation in math courses

a parent interested in encouraging your daughter in math and science

a community youth leader seeking good math and science activities for out-of-school groups

an employer looking for women with scientific or technical training

a woman resuming your education or career and lacking prerequisite math training

THIS IS WHAT THE NETWORK HAS FOR YOU

PROGRAMS—to increase girls' and women's confidence and competence in doing mathematics.

STRATEGIES—to motivate girls and women toward new career options

CONTACTS—for students and employers with women working in nontraditional occupations

INSERVICE—for teachers and counselors in elementary and secondary schools.

MATERIALS—for teachers and parents to encourage students in math-based fields of work and study.

THE MATH/SCIENCE NETWORK represents 500 educators, scientists, engineers, administrators, and community people from 30 colleges and universities, 20 school districts, and numerous corporations, government agencies, and foundations. **NETWORK** members have been working cooperatively since 1974 to develop and conduct innovative math and science education programs for girls and women. The **NETWORK** has conducted 20 conferences, reaching 7,000 secondary, college, and professional women.

THE MATH/SCIENCE RESOURCE CENTER

is the main clearinghouse and switchboard of the **NETWORK**. Inquiries should be directed to:

Joanne Koltnow, Coordinator
Math/Science Resource Center
Mills College
Oakland, California 94613
(415) 635-5074 or 635-9271

PROJECT DIRECTORS:

Lenore Blum
Director of Collegiate Programs
Mathematics and Computer Science
Mills College
Oakland, California 94613
(415) 632-2700, ext. 347

Nancy Kreinberg
Director of Secondary Programs
Lawrence Hall of Science
University of California
Berkeley, California 94720
(415) 642-1823

ADVISORY COMMITTEE*

Dorothy L. Bernstein, President, Mathematical Association of America

Janet Welsh Brown, Executive Director, Environmental Defense Fund

Yolanda Scott George, Biologist and Affirmative Action Coordinator, Lawrence Livermore Laboratory

Shirley Hill, President, National Council of Teachers of Mathematics

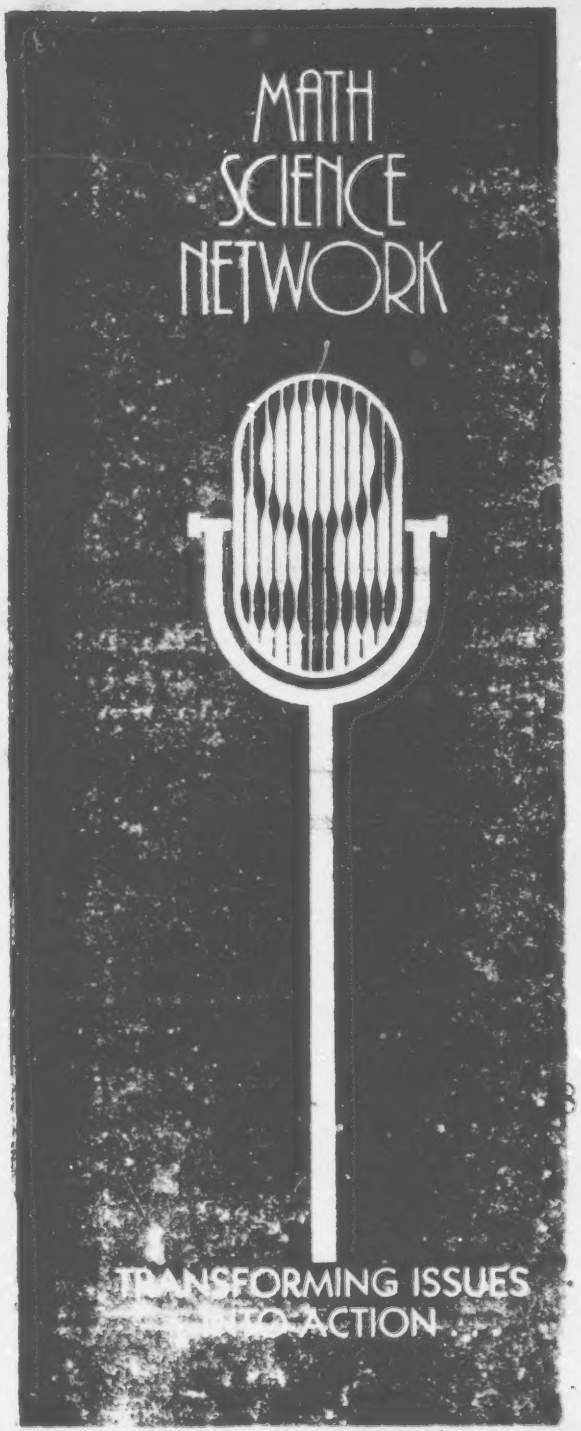
Madeline Mixer, Regional Director, Women's Bureau, U.S. Department of Labor

Bernice Sandler, Director, Project on the Status and Education of Women

Joy R. Simonson, Executive Director, National Advisory Council on Women's Educational Programs

Betty M. Vetter, Executive Director, Scientific Manpower Commission

*Organizations listed for identification purposes only



ATTACHMENT IV

Banesh Hoffmann's letter regarding flaws in a math competency test.

THE NEW YORK TIMES, SATURDAY, SEPTEMBER 22, 1979

Letters

Flaws in a Math Competency Test

To the Editor:

In a report in The Times on Sept. 14 about the high rate of failure on the mathematics competency test one reads, "In one question students were asked to use the Pythagorean theorem — with which they were provided — to determine the length of the hypotenuse of a right triangle whose legs are 3 and 4. They were given four multiple-choice possibilities: (a) 5, (b) 7, (c) 12 and (d) 25. The correct answer is 5."

Students who picked the sum or product of 3 and 4 — choice b or c — certainly gave no evidence of understanding relevant concepts. But what of students who picked choice (d) — the sum of the squares of 3 and 4? In all likelihood many of them well understood the relevant concepts and under normal circumstances would be able to obtain the correct result; but under the stress of the occasion, and seeing the enticing choice (d) 25, they forgot to take the square root of the 25.

Choice (d) was included as a deliberate trap that, unlike choices (b) and (c), was meant to ensnare able students. But the purpose of the test is, or should be, to measure mathematical competence and not behavior under stress. Inclusion of choice (d) conflicts with that purpose.

What of the students who picked the correct choice? Did they all understand the relevant concepts? Not necessarily, by any means. This is because the question can be correctly answered without any understanding of the Pythagorean theorem: a student need merely sketch a right triangle with legs 3 and 4 and note its hypotenuse.

All this being so, we may well wonder, among other things, why choice (a) earns full credit while choice (d) is treated as on a par with choices (b) and (c).

BANESH HOFFMANN

Flushing, N.Y., Sept. 15, 1979

The writer is a professor emeritus of mathematics at Queens College.

ATTACHMENT V

Summary of
Federal Court Decision in Case of
Larry P. v. Riles

IN THE UNITED STATES DISTRICT COURT

FOR THE NORTHERN DISTRICT OF CALIFORNIA

LARRY P., by his Guardian
ad Litem, LUCILLE P., et al.,

Plaintiffs,

No. C-71-2270 RFP

v.

WILSON RILES, Superintendent
of Public Instruction for the
State of California, et al.,

Defendents.

O P I N I O N

ORIGINAL
F I L E D

OCT 10, 79

WILLIAM L. WHITAKER
CLERK, U. S. DIST. COURT
SAN FRANCISCO

Plaintiffs, representing the class of black children in California who have been or in the future will be wrongly placed and maintained in special classes for the "educable mentally retarded" ("E.M.R.")[1/] challenge the placement process for those classes and particularly certain uses of standardized individual[2/] intelligence ("I.Q.") tests in California.[3/] They contend that the I.Q. tests in their present form are biased and that defendants have discriminated against black children by using those tests. The tests allegedly result in the misplacement of black children in special classes that doom them to stigma, inadequate education, and failure to develop the skills necessary to productive success in our society.[4/] Black children represent only 10 percent of the present general student population in California, but provide some 25 percent of the population enrolled in E.M.R. classes.

These testing and placement problems arise in a setting of educational failure. California's schools have been unable to meet the educational needs of disadvantaged minorities such as the black children who brought this case.[5/] As a result, poor minority children tend not only to start out behind their white, middle-class counterparts, but also tend to fall increasingly farther behind after exposure to the public school system. Defendent Wilson Riles stated in 1969, before he became Superintendent of Public Instruction, that disadvantaged black children, on the average, learn only about .7 as much as middle-class white children in any given year.[6/] Other, more recent data confirmed that analysis. Black children fall increasingly behind to the point that it is not unusual for high school students to be reading at the third grade level and performing at only the fourth grade level in mathematics.[6a/] The disproportionate

placement of black children into E.M.R. classes is but one aspect of this troublesome situation.

Courts cannot solve our educational problems, but they played a part in the incremental effort to improve those aspects of our educational systems that effectively deny minorities an equal opportunity to succeed. In particular, the phenomenon of special education such as that for the "mentally retarded" has not yet been subjected to much judicial scrutiny.[7/] We have been forced in this case to enter that complicated area, and it raises special problems for court intervention.

A principal focus of this litigation is on testing -- on the use of individual I.Q. tests -- to classify black children and assign them to E.M.R. classes. Much of the more than 10,000-page transcript of the trial represents detailed expert testimony about these tests. The court has necessarily been drawn into the emotionally charged debate about the nature of "intelligence" and its basis in "genes" or the "environment." This debate, which finds reknowned experts disagreeing sharply, obviously cannot be resolved by judicial decree. Despite these problems, however, court intervention has been necessary. The history of this litigation has demonstrated the failure of legislators and administrative agencies to confront problems that clearly had to be faced, and it has revealed an all to typical willingness either to do nothing or to pass on issues to the courts.

Fortunately, the "scientific controversy" surrounding the I.Q. tests has not materialized to the extent that might have been expected. The experts have tended to agree about what I.Q. tests can and cannot do, even if they disagree about the utility of I.Q. testing for E.M.R. placement. Our decision,

therefore, rests more on a consensus than on the testimony of any one line of experts. Given that consensus, coupled with the other factors present in this case, there is no choice but to invalidate California's present system of classification of black children for E.M.R. classes. The bases for this ruling, both statuatory and constitutional, will be explained in detail below, but it may be helpful to summarize them briefly at the outset before proceeding to the main body of the opinion.

SUMMARY

This court finds in favor of plaintiffs, the class of black children who have been or in the future will be wrongly placed or maintained in special classes for the educable mentally retarded, on plaintiffs' statutory and state and federal constitutional claims. In violation of Title VI of the Civil Rights Act of 1964, the Rehabilitation Act of 1973, and the Education for All Handicapped Children Act of 1975, defendants have utilized standardized intelligence tests that are racially and culturally biased, have a discriminatory impact against black children, and have not been validated for the purpose of essentially permanent placements of black children into educationally dead-end, isolated, and stigmatizing classes for the so-called educable mentally retarded. Further, these federal laws have been violated by defendants' general use of placement mechanisms that, taken together, have not been validated and result in a large overrepresentation of black children in the special E.M.R. classes.

Defendents' conduct additionally has violated both state and federal constitutional guarantees of the equal protection of the laws. The unjustified toleration of disproportionate enrollments of black children in E.M.R.

classes, and the use of placement mechanisms, particularly the I.Q. tests, that perpetuate those disproportions, provide a sufficient basis for relief under the California Constitution. And under the federal Constitution, especially as interpreted by the Ninth Circuit Court of Appeals, it appears that the same result is dictated.

Moreover, there is another basis for the federal constitutional ruling. Defendents' conduct, in connection with the history of I.Q. testing and special education in California, reveals an unlawful segregative intent. This was not necessarily to hurt black children, but it was an intent to assign a grossly disproportionate number of black children to the special E.M.R. classes, and it was manifested, inter alia, in the use of unvalidated and racially and culturally biased placement criteria. This intent, consistent only with an impermissible and unsupportable assumption of a higher incidence of mental retardation among blacks, cannot be allowed in the face of the constitutional prohibition of racial discrimination.

The precise reasons for these findings, expressed here in simplified form, and the type of remedy that will be necessary, including a continuation of the present injunction against I.Q. testing for the placement of black children into E.M.R. classes, are discussed in the body of this opinion.

ATTACHMENT VI
Description of NIE - sponsored
Study of School District Uses of Evaluation and Testing Information

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
NATIONAL INSTITUTE OF EDUCATION
WASHINGTON, D.C. 20208

THE STUDY OF SCHOOL DISTRICT USES OF EVALUATION AND TESTING INFORMATION

In late September of 1979, the National Institute of Education awarded a two-year contract to the Huron Institute to study school district uses of evaluation and testing information. The study consists of two interrelated activities: (1) the in-depth case study of 12 school districts that have successfully tied evaluation and testing information to instructional renewal and managerial reform. Less intensive study will be made of an additional 10 districts which have promising practices. (2) the development of technical assistance strategies to assist other school districts in better using evaluation and testing information.

Jointly funded by the National Institute of Education and the U.S. Office of Education, the study is designed to tie directly the results of research to the creation of program initiatives. The study reflects previous work by both agencies in this area.

The Huron Institute is now searching for school districts who have had successful experiences with the uses of testing and evaluation information. We would appreciate your suggestions for school districts to include in this study. The value of the study rests heavily on the extent to which school districts selected can shed light on this important topic.

To convey recommendations about school districts for possible inclusion and to obtain more information about the study, please contact:

Dr. Mary Kennedy
The Huron Institute
123 Mt. Auburn St.
Cambridge, Mass
(617) 491-5450

At NIE, contact:

Ms. Mary Ann Millsap
Program on Testing, Assessment
and Evaluation
National Institute of Education
Washington, D.C. 20208
(202) 254-6271

ATTACHMENT VII
Letter from Mary Ann Millsap

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
NATIONAL INSTITUTE OF EDUCATION
WASHINGTON, D.C. 20208

October 29, 1979

Dear Colleague:

The Methodology section of the 1980 Teaching and Learning Grants Announcement strongly encourages school districts to submit grant applications on issues related to the uses of testing and evaluation information. As part of the Institute's mission to improve local educational practice, the Program on Testing, Assessment and Evaluation developed this grants program primarily around issues related to uses of testing and evaluation information in school districts. Pages 20-23 of the announcement describe this focus in greater detail.

Grant applications are due no later than January 21, 1980, and we anticipate grant awards in May 1980. From 15 to 30 grants of varying sizes will be awarded. We hope this grants mechanism will help us explore evaluation and test use and assist us in building stronger networks in the field.

This grants program complements other research activities the Institute is undertaking, including several studies of test use as well as a two-year study, jointly funded by the U.S. Office of Education, of school districts that have successfully tied evaluation and testing information to instructional renewal and managerial reform. Technical assistance strategies to assist school districts in better use of such information will also be developed through this contract.

This grants announcement is the first major effort to focus grant awards on school districts. For us to ensure creating and sustaining this effort, we are soliciting a large pool of high quality proposals from which to draw these awards. We strongly urge you either to apply for a grant or to encourage others to do so. For further information, please call either Corinne Scott or myself (both on 202-254-6271) or to write us c/o The Program on Testing, Assessment and Evaluation, National Institute of Education, 1200--19th Street, N.W., Washington, D.C. 20208.

Sincerely,

Mary Ann Millsap

Mary Ann Millsap
Senior Associate

ATTACHMENT VIII
Sample materials destributed by Amity Buxton

OAKLAND UNIFIED SCHOOL DISTRICT
Department of Research and Evaluation
Division of Learning

ONGOING DOCUMENTATION OF STUDENT PROGRESS: THIRD AND FIFTH GRADE

Background and use

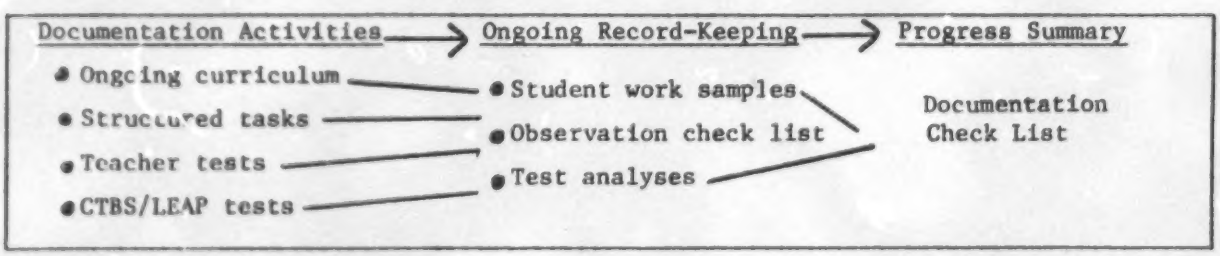
This packet of documentation activities was selected and designed by a group of teachers who worked during the summer of 1979 with staff members from the Division of Learning and the Research Department. It was prepared specifically in response to the newly-adopted Board policy concerning pupil promotion, retention and acceleration and, at the fifth grade level, State requirements for early assessment of proficiencies.

While the new Board policy concerns the establishment of standards and procedures for identifying students who may need to be retained at all grade levels, specific retention criteria were adopted for grades three and five. One of those criteria is the students' demonstrated mastery of 70% of the key grade level learning expectations.

In a survey conducted by the Division of Learning in Spring 1979, a cross-section of district teachers identified which learning expectations are "key" to the continued success of students as they progress through the elementary school curriculum. All of these expectations can be taught, assessed and documented through ongoing curriculum. Some can be tested by formal paper/pencil tests. Others require supplementary assessment activities through which student achievement can be measured against standard criteria by teacher observation or analysis of samples of student work. This documentation covers them all.

The accompanying documentation activities for fifth grade cover all of the key expectations as well as the proficiency skills included therein. By using these activities, fifth grade teachers will document the required areas for grades 4-6 early proficiency assessment and key learning expectations for promotion/retention decisions.

In the process of developing these documentation procedures. a high priority was placed on integrating the activities with existing materials and ongoing classroom procedures. The following diagram summarizes the structure which underlies the use of the suggested activities for keeping track of student progress and achievement:



For the purpose of summarizing student progress in relation to the key expectations (and early proficiency skills) we have devised the DOCUMENTATION CHECK LIST included in this manual. This check list will replace the mathematics and reading profile cards used previously at grades three and five.

Documentation Activities -

We hope the sample activities will become part of your ongoing curriculum, although they may be used as a separate checkup process as well. The documentation activity cards are numbered according to the mathematics and reading/language learning expectations lists. The documentation activities include individual, small group, and large group activities in all subject areas and utilize a variety of resources such as regular class work, teacher dictation, chalkboard presentations, learning centers, worksheets, manipulatives, discussions, commercial kits or lessons, textbooks, reference books, the school library, and reports.

The activities are designed to enable students to demonstrate their mastery of the key expectations and early proficiency skills in their ongoing class activities through written work which you can keep as assessment data or performance which you can observe and check off on observation check lists. The activity cards specify the criteria to be used in deciding if the students have mastered the skills involved. The cards also indicate what you will need to keep for your cumulative records.

Each card in the packet of documentation activities follows the same format, in which the key expectation (and early proficiency skill at the fifth grade level), sample activities, criteria, and data to be kept for documentation are all indicated, as follows:

OUSD-KEY EXPECTATIONS DOCUMENTATION — GRADE-TOPIC-CARD # (Key Expectation #)

Key expectation: Statement of learning expectation.

(For fifth grade only, Early proficiency skill: Statement of early proficiency skill.)

To get to this → Statement of objective.

Try this → Sample activity.

Look for this → Standard criteria.

And keep this → Data to be kept for documenting student progress, such as:

Work samples (one sample may serve to document more than one learning expectation.)

Observation check list of student performance

Test analyses (Your own and LEAP and CTBS)

(RELATED DOCUMENTATION ACTIVITIES)

For some expectations, formal assessment data may be obtained from the CTBS or LEAP tests. In these cases, the appropriate test item numbers are noted on the card.

Each card is coded in the top right corner with the grade level, the area (MATH or READING/LANGUAGE), the card number, and the learning expectation number to assist you in locating additional informal assessment and instructional activities in the curriculum guides. In the lower right hand corner are related documentation activities. These are noted for your use because of their implications for ongoing diagnosis and individualization of your curriculum.

Procedures

1. Read through the cards to see what kinds of activities are described and how they can fit into your curriculum. You may rearrange them in any order which is most appropriate for your class. Note also the cross-referencing of many of the activities (listed in the manual and noted on the cards) to see how they reinforce one another.

2. Prepare individual work folders for keeping your students' work samples.

3. If you don't have your own, make copies of the sample observation check lists which are included in this manual so you can begin to use them in your daily classwork.

4. If you need them, feel free to copy the sample work activities listed on many of the cards.

5. Look at the DOCUMENTATION CHECK LIST to see where you will record each students' mastery of each key expectation and proficiency.

Recordkeeping and Progress Summary

There is such a strong relationship between the key expectations, the documentation activities, the district's curriculum guides and the student profile charts that the documentation activities will be useful for obtaining basic information for updating the student progress records in mathematics and reading/language. To avoid duplication, the DOCUMENTATION CHECK LIST will serve as the students' profile record at grades three and five. (Because the documentation covers most of the 4-6 early proficiency skills, the DOCUMENTATION CHECK LIST will in turn be useful for summarizing early proficiency progress at the fifth grade level.)

The student work samples which you keep (including math worksheets, daily journals, class work, class tests, etc.) and your observations of student performance will provide an excellent "data bank" which can document student progress. As you summarize information from the work samples, observations and tests and transfer it onto the DOCUMENTATION CHECK LIST throughout the year, you will be laying the groundwork for making careful decisions about promotion, retention or acceleration for each student.

We welcome your suggestions as you work with these activities. If you want help in implementing them, call Amity Buxton (Teacher Shelter, 836-8316) or Joe Trimble (Research Department, 836-8257). Later in the year we will be requesting your feedback on the activities and process for incorporation into the District's formal assessment procedures for next year.

When and how to use documentation

- to satisfy state mandate for early assessment (grade 4-6) of graduation proficiency skills - per Assembly Bill 65
- to keep track of individual student's on-going progress
- to identify areas of student growth needing special help and alternative teaching/learning techniques
- to confer with parents
- to confer with other professionals (principal, psychologist,etc.)
- to inform your sub about where students are
- to decide whether a child can be promoted or retained
- to inform the next teacher of each student's progress when the child is transferred, promoted,or retained.

If you want help or answers

First - *call* Your school site curriculum assistant, assistant principal, or resource teacher

Then one of the following:

District Mathematics Team:	Charles Barker	8303
	Bill Self	8271
District Reading/Language Team	Yetive Bradley	8303
Staff Development Team:	Carol Worthington	8292
Teacher Shelter Staff:	Amity Buxton	8316
Research Department:	Joe Trimble	8257

And *go to* Teacher Shelter - Tuesday 9-8 p.m.
Wednesday 9-6 p.m.
Thursday 9-6 p.m.

GRADE THREE

SAMPLE LIST OF READING/LANGUAGE AND MATHEMATICS KEY EXPECTATIONS

Cross References among Key Expectations, Documentation Activities and CTBS Items

Key Expectation		Related Key Expectations and Documentation Activities		Related CTBS Test Item
Card #	Key Expectation	Card #	Key Expectation	Test Item
	READING/LANGUAGE SAMPLE ITEMS			
4	Give brief talk about an experience using descriptive details	6	speak clearly and adjust speed, volume, pitch	
		7	express ideas and feelings in a variety of situations	
		17	select main ideas and details	
		3	ask relevant questions about information given orally	
17	Select main ideas and details	8	write paragraph with beginning, middle, end	2 1,3,5
		10	recognize and use comma, apostrophe (in addition to period and question mark)	14-15
		11	use additional capitals in title, abbreviations (beginning of sentence and I)	34
		12	recognize and form letters in cursive writing	
	MATHEMATICS SAMPLE ITEMS			
10	Add and subtract 2 and 3 and 4 place numbers with/without regrouping	4	write numbers through 1000	6: 3,5,7
		21	identify value coins and bills, make change	13-19 21-23
11	Add 2-place column with 3 or 4 addends	4	write numbers through 1000	6: 8,9
		21	identify value coins and bills, make change	
23 *	Write, analyze and solve work problems	6	write number words one to twenty	7: 13,20
		16	use symbols +, -, x, ÷, <, and >	
		24	change word problems to number	27
24	Change word problems to number problems	10	add and subtract 2 and 3 and 4 place number with/without regrouping	
		11	add 2 place column with 3 or 4 addends	
		6	write number words one to twenty	
		23	write, analyze and solve word	

* THE CARD EXPLAINING THIS TOPIC IS INCLUDED IN THE SAMPLE PACKET

OUSD - KEY EXPECTATIONS DOCUMENTATION

GR. 3 - MATH
(Problem Solving/App.-1) 23

Key expectation: Write, analyze, and solve word problems.

To get to this ⟶ Given directions to devise a word problem from a practical situation, the student can write a problem and other students can analyze and solve the problem.

Try this ⟶

Each student devises a word problem about a school situation (ex. Each class has three balls for 30 students. How can the balls be divided so that each child has an equal time with the ball during recess?) Then each student has to analyze and solve the problems written by 5 other students (selected as solvable by the teacher).

Look for this ⟶
- identification of a problem; statement of a question
- making a numerical problem out of the situation
- putting 4 out of 5 problems into mathematical figures
- correct solution of 4 out of 5 of the problems

And keep this ⟶ Student work (Note: work samples can serve to document more than one expectation)

AND

Give this ⟶ CTBS Test 7

Look for this ⟶ Items # 13-26, 27-29

And keep this ⟶ Item analysis from Research

* See also Math Card 6:Counting
Math card 16:Operations 3
Math card 24:Prob.Solv.Appli.2

ATTACHMENT IX

INITIAL IMPACT REPORT OF THE ADMISSION TESTING LEGISLATION
By State Department of Education of New York, Oct. 16, 1979.

SUMMARY

Twenty-six postsecondary school admission testing programs used in New York State have been identified by the test agencies. Based on very tentative information available by October 15, about twenty of these programs may be withdrawn from use in the State. Postsecondary institutions will be surveyed within the next few weeks to help determine the current use of these test programs and to identify other testing programs being used by the institutions for admission selection purposes. It is likely that other tests, not designed for admissions purposes, are actually being used in this way by the institutions.

Five of the six programs expected to continue in the State are published by the Educational Testing Service for various sponsors including the College Board. The sixth is the ACT Assessment Program sponsored by The American College Testing Program. The sponsors of these six test programs have all indicated their desire to continue testing, however, they all expect to reduce services. The number of regular administrations is expected to be reduced somewhat and special testing will be curtailed sharply. This step is likely to have the greatest effect on the physically handicapped and sabbath observers.

About half of the twenty test programs which may be withdrawn have very limited usage in the State and their elimination may not produce a severe impact. The other half, however, include several tests used for admission to highly competitive health related programs such as medicine and dentistry. Also to be withdrawn are tests such as the Miller Analogies Test which is used by many graduate level programs in the State. Withdrawal of these tests is expected to have the greatest impact on applicants with prior academic backgrounds which are questionable or unfamiliar to the institution.